



Международный журнал информационных технологий и энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.9

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ПРОГНОЗИРОВАНИЯ ПОСЕЩАЕМОСТИ ВЕБ-САЙТОВ

¹Бондаренко О.С., ²Кипилова А., ³Смирнов А.А.

ФГАОУ ВО "НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ (ИТМО)", Санкт-Петербург, Россия (197101, город Санкт-Петербург, Кронверкский пр-кт, д. 49 литер а), e-mail: ¹rancerenly@gmail.com, ²akipilova@gmail.com, ³smirnov.andrew.1999@yandex.ru

В контексте информационного общества, эффективное управление веб-трафиком становится критическим для функционирования веб-сайтов, служащих основными платформами для коммуникации и коммерции. Проблемы, связанные с колебаниями трафика, такие как сбои и замедление загрузки, могут негативно сказаться на бизнес-показателях, особенно в периоды пиковой активности пользователей. Использование аналитических инструментов, включая Google Analytics и Grafana, обеспечивает сбор и анализ данных о посещаемости, что способствует прогнозированию тенденций и оптимизации маркетинговых стратегий. Методы прогнозирования, такие как ARIMA и LSTM, основанные на временных рядах, позволяют с высокой точностью предсказывать будущий трафик, что является ключевым для адаптации ресурсов веб-сайтов к изменяющимся условиям. Точные прогнозы трафика предоставляют необходимые данные для управления ресурсами, оптимизации нагрузки и предотвращения технических проблем, тем самым улучшая пользовательский опыт и поддерживая стабильность бизнес-операций.

Ключевые слова: Веб-трафик, аналитические инструменты, прогнозирование трафика, пользовательский опыт, веб-технологии.

COMPARATIVE ANALYSIS OF METHODS FOR PREDICTING WEBSITE TRAFFIC

¹Bondarenko O.S., ²Kipilova A., ³Smirnov A.A.

NATIONAL RESEARCH UNIVERSITY OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS (ITMO), St. Petersburg, Russia (197101, St. Petersburg, Kronverkskiy pr-kt, 49, lit.a), e-mail: ¹rancerenly@gmail.com, ²akipilova@gmail.com, ³smirnov.andrew.1999@yandex.ru

In today's information society, managing web traffic effectively is crucial for the success of websites that serve as the primary platforms for communication and commerce. Issues related to traffic fluctuation, such as sudden spikes and slow downloads, can significantly impact business performance, particularly during peak user activity periods. The use of analytical tools like Google Analytics and Grafana enables the collection and analysis of traffic data. This data helps predict trends and optimize marketing strategies, ensuring a smooth user experience. Forecasting techniques such as ARIMA (AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory) models, based on historical traffic patterns, allow for highly accurate predictions of future traffic. These predictions are essential for adapting website resources to changing conditions and preventing technical problems. Accurate traffic forecasts provide vital data for managing resources, optimizing load, and preventing downtime. This, in turn, enhances the user experience and maintains the stability of business operations, ultimately contributing to long-term success.

Keywords: Web-traffic, analytical tools, traffic forecasting, user experience, web-technologies.

Современное информационное общество охватывает все сферы человеческой жизни, и веб-сайты становятся ключевым инструментом для обмена информацией, предоставления услуг и ведения бизнеса.

Способность справляться с колебаниями трафика является решающим фактором успеха любой организации. Увеличение трафика веб-сайта может создать несколько проблем, поскольку многие люди сталкивались со сбоем веб-сайта или медленной скоростью загрузки из-за большого трафика. Например, во время праздников интернет-магазины могут выйти из строя из-за наплыва пользователей, пытающихся войти на платформу, превышающую ее возможности. Такие сценарии могут привести к снижению рейтингов пользователей и заставить их переключиться на другие веб-сайты, тем самым снижая бизнес организации. Веб-сайты должны отображать веб-страницу посещения для распределения компьютерных ресурсов, прогнозирования будущих доходов и прогнозирования будущего роста рекламы. Чтобы определить популярность определенных разделов или страниц и выявить тенденции, веб-сайты отслеживают как входящий, так и исходящий трафик. Это позволяет им анализировать такие данные, как частота посещения определенной страницы посетителями из определенной страны.

Анализ инструментов для сбора информации посещаемости веб-сайта

Прогнозный анализ посещаемости веб-сайта становится ключевым элементом для компаний, основной деятельностью которых является онлайн-торговля. С увеличением числа интернет-пользователей каждый год наблюдается значительный рост посещаемости веб-сайтов. Согласно данным различных источников, прирост трафика на популярных веб-ресурсах составляет в среднем 10% ежегодно. Социальные сети стали одними из наиболее посещаемых рекламных площадок в онлайн-пространстве. Данные ресурсы служат платформами, где ежедневно сотни и тысячи пользователей просматривают и взаимодействуют с рекламой. Способность адекватно предсказывать и анализировать рост трафика играет критическую роль в эффективном управлении онлайн-торговлей и формировании успешных маркетинговых стратегий. [1]

Статистика в сфере интернета представляет собой информацию о том, как часто и кем посещается веб-ресурс. Борисевич А. С. [2] под понятием "посещаемость" понимает количество загрузок страниц сайта пользователем сети. Эти данные особенно важны для владельцев веб-ресурсов. Имея доступ к статистике, они могут делать выводы о результативности рекламных кампаний, интересе пользователей к контенту сайта, технических характеристиках посетителей и многом другом. Это особенно актуально, учитывая, что многие веб-сайты служат рекламными площадками.

Веб-аналитика автоматизирует процесс сбора и систематизации данных, облегчая анализ и принятие решений в цифровом маркетинге. С ее помощью измеряется поток кликов, наблюдаются источники посещаемости и поведение посетителей. [3]

Информация о трафике собирается на аккаунте рекламодателя и анализируется с помощью отчетов. Эти отчеты содержат различные метрики, такие как количество посетителей на сайте, количество кликов по рекламе, глубина просмотра, стоимость клика и другие. Полученные данные могут быть экспортированы в формате .xml для дальнейшего использования. [1]

Google Analytics является ведущим инструментом веб-аналитики с множеством отслеживаемых параметров, таких как процент отказов, количество веб-сессий, средняя продолжительность сессии и другие. [3]

Одним из основных средств анализа данных для прогнозирования динамики нагрузки является визуализация, которая помогает представить численные статистические данные в удобной форме через графику, диаграммы и гистограммы. Для этого используются специализированные программы. Примером такого инструмента для визуализации статистических данных с веб-серверов является пакет Grafana. [Ошибка! Источник ссылки не найден.]

Анализ инструментов для прогнозирования веб-трафика

Прогнозирование веб-трафика является активной областью исследований, в которых применяются различные методы, такие как вейвлет-анализ, нейронные сети и генетические алгоритмы. Все методы используют прогнозирование временных рядов. [5]

Временные ряды — это статистические данные, расположенные и представленные в хронологическом порядке в течение определенного периода времени, как объяснили Casado-Vara R. и др [6].

Для нелинейного прогнозирования используются два подхода, которые рассмотрели ученые в исследовании "Web traffic time series forecasting using ARIMA and LSTM RNN" [7]. Они заключаются в следующем:

- ARIMA Модель,
- LSTM Модель.

Для оценки эффективности модели используются количественные методы, такие как коэффициент детерминации (R^2) и средняя абсолютная процентная ошибка (MAPE). MAPE, также известная как среднее абсолютное процентное отклонение (MAPD), измеряет точность прогнозирования. Коэффициент детерминации (R^2) показывает, насколько хорошо модель объясняет дисперсию зависимых переменных. Высокая точность прогнозирования отражается в значениях $R^2 = 0,9996$ и $MAPE = 0,0024$. [0]

Модель ARIMA (Autoregressive Integrated Moving Average) — это статистический инструмент для анализа и прогнозирования временных рядов данных. Она явно учитывает набор стандартных структур в данных временных рядов и, таким образом, предоставляет простой, но мощный метод для составления искусственных прогнозов временных рядов.

Модель включает в себя дифференцирование (интегрирование) исходных данных временного ряда. Дифференцирование временных рядов означает формирование нового временного ряда путем вычитания предыдущего наблюдения из текущего времени. Смысл этого заключается в устранении определенных тенденций, таких как сезонность, тренды или непоследовательная дисперсия в данных временного ряда. [7]

Параметры модели ARIMA определяются следующим образом:

- **p**: Порядок лага, представляющий собой количество наблюдений лага, включенных в модель.
- **d**: Степень дифференцирования, обозначающая количество раз, когда исходные наблюдения подвергаются дифференцированию.

- **q**: порядок скользящего среднего, указывающий на размер окна скользящего среднего. [0]

Строится линейная регрессионная модель, включающая указанное количество и тип членов, а данные подготавливаются с помощью дифференцирования, чтобы сделать их стационарными, то есть убрать трендовые и сезонные структуры, которые негативно влияют на регрессионную модель. [10]

Интересно, что любой из этих параметров может быть установлен в 0. Такие конфигурации позволяют модели ARIMA имитировать функции более простых моделей, таких как ARMA, AR, I или MA. [7]

Длительная кратковременная память (LSTM = Long Short-Term Memory) — это тип рекуррентной нейронной сети (RNN), которая широко используется в прогнозировании сетевого трафика. В отличие от традиционных РНС, которые могут страдать от проблема исчезающего градиента, LSTM разработаны для запоминания долгосрочных зависимостей и избежания проблемы исчезающих градиентов. Это делает их хорошо подходящими для решения задач прогнозирования временных рядов таких как прогнозирование сетевого трафика. [11]

В прогнозировании сетевого трафика LSTM моделируют временные зависимости между данными, обучаясь на исторических данных и предсказывая будущий трафик. LSTM улавливают сложные закономерности, включая ежедневные и еженедельные, и делают точные прогнозы даже при наличии шума или недостающих данных. LSTM используются в различных приложениях для прогнозирования сетевого трафика, включая прогнозирование трафика в центрах обработки данных, облачных вычислительных средах и беспроводных сетях. [11]

Casado-Vara R. и др. [6] исследовали прогнозирование мирового веб-трафика Википедии по выбранным темам, используя данные, собранные собственным скрепером Wikipedia. Они охватили все языки по списку ISO 639-1, что расширило возможности модели. Исследование фокусировалось на использовании модели LSTM для анализа временных рядов.

Авторы исследования "Web traffic time series forecasting using LSTM neural networks with distributed asynchronous training" [6] осознали ограничения LSTM в прогнозировании аномалий без учета сезонности и тренда, однако отметили, что их модель немного превосходит некоторые современные модели для прогнозирования веб-трафика. Для будущих исследований предлагается изучение влияния новостей из социальных сетей на веб-трафик.

В статье "Web traffic time series forecasting using ARIMA and LSTM RNN" [7] представлена методика прогнозирования временных рядов веб-трафика с использованием ARIMA (Auto-Regressive Integrated Moving Average) и LSTM RNN (Long Short Term Memory Recurrent Neural Network). Авторы стремятся решить проблемы, возникающие в связи с ростом веб-трафика на сайтах, что может привести к таким проблемам, как перебои в работе серверов и медленное время загрузки.

Дискретное вейвлет-преобразование (DWT). DWT применяется к вектору данных, разделяя его на точные и приближенные коэффициенты с помощью последовательности фильтров низких и высоких частот. В результате формируются низкочастотные и высокочастотные компоненты, которые восстанавливаются с помощью

Shelatkar T. и др. [7] использовали API Википедии для создания временного ряда ежедневных посещений страниц. Данные обработали с помощью дискретного вейвлет-преобразования (DWT), разделив их на низкочастотные и высокочастотные компоненты. Для прогнозирования линейных и нелинейных компонент применили модели ARIMA и LSTM RNN соответственно, затем объединили прогнозы с помощью обратного вейвлет-преобразования (iDWT). Это улучшило точность прогнозов.

Анализ источников показал, что экспериментальные методы на основе временных рядов широко используются для прогнозирования посещаемости веб-сайтов, позволяя разработчикам предвидеть изменения в структуре посещаемости сайтов и соответствующим образом корректировать их работу. Эти методы опираются на исторические данные о посещаемости сайта для получения прогнозов будущих объемов трафика, что позволяет разработчикам выявлять потенциальные пики посещаемости и вносить необходимые коррективы для оптимизации работы сайта.

Этот подход позволяет лучше улавливать сложности в динамике веб-трафика, включая сезонные изменения, колебания и влияние внешних факторов. Полученные результаты обеспечивают практические рекомендации для сетевых администраторов, маркетологов и руководителей веб-проектов. Точные прогнозы, достигнутые с использованием различных моделей, могут помочь оптимизировать управление ресурсами, распределять нагрузку в периоды пикового трафика и предотвращать проблемы с производительностью сайтов.

Список литературы

1. Фомичев С. С., Терешков А. А. Прогнозирование посещаемости веб-сайтов из необработанных данных с использованием сетей LSTM // *Modern Science*. 2020. № 5–3. С. 591–597.
2. Борисевич А. С. Математическое прогнозирование в интернет-статистике. – 2003.
3. Ларионов В. С., Дунин И. В. АНАЛИЗ РАБОТЫ ВЕБ-СЕРВЕРОВ С ИСПОЛЬЗОВАНИЕМ ПАКЕТА GRAFANA // *Форум молодых ученых*. 2017. №11.
4. Петлури Н., Аль-Масри Э. Прогнозирование веб-трафика страниц Википедии // *Международная конференция IEEE 2018 по большим данным (Big Data)*. – IEEE, 2018. – С. 5427-5429.
5. Шелаткар Т. и др. Прогнозирование временных рядов веб-трафика с использованием ARIMA и LSTM RNN // *ITM Web of Conferences*. – EDP Sciences, 2020. – Т. 32. – С. 03017.
6. Сикка Д., Кумар К. Н., С. В. Прогнозирование временных рядов посещаемости веб-сайта с использованием регрессионного машинного обучения // *12-я Международная конференция IEEE по системам связи и сетевым технологиям (CSNT)*, 2023. – IEEE, 2023. – С. 246-250.
7. Чжоу К. и др. Сравнительное исследование временных рядов прогнозирования веб-трафика на основе статистической модели и генеративной состязательной модели // *Системы, основанные на знаниях*. – 2021. – Т. 213. – С. 106467.
8. Матлакунта Р. Р. Прогнозирование веб-трафика для онлайн-рекламы : дис.... – Оклэндский технологический университет, 2011.
9. елс Дж. Прогнозирование веб-трафика с использованием моделей авторегрессии, LSTM и XGBoost временных рядов // *Прикладные исследования в области искусственного интеллекта и облачных вычислений*. – 2020. – Т. 3. – №. 1. – С. 1-15.

10. Яо С., Ху С., Сун М. Прогнозирование веб-трафика на основе вейвлетов и нейронных сетей // 6-й Всемирный конгресс по интеллектуальному управлению и автоматизации 2006. – IEEE, 2006. – Т. 1. – С. 4026-4028.
11. Джун С. П., Сон Т. Э., Парк Х. В. Прогнозирование по аналогии с использованием веб-поискового трафика // Технологическое прогнозирование и социальные изменения. – 2017. – Т. 115. – С. 37-51.

References

1. Fomichev S. S., Tereshkov A. A. Forecasting website traffic from raw data using LSTM networks // Modern Science. 2020. No. 5-3. pp. 591-597.
 2. Borisevich A. S. Mathematical forecasting in Internet statistics. – 2003.
 3. Larionov V. S., Dunin I. V. ANALYSIS OF THE WORK OF WEB SERVERS USING THE GRAFANA PACKAGE // Forum of Young Scientists. 2017. №11.
 4. Petluri N., Al-Masri E. Web traffic prediction of wikipedia pages // 2018 IEEE International Conference on Big Data (Big Data). – IEEE, 2018. – pp. 5427-5429.
 5. Shelatkar T. et al. Web traffic time series forecasting using ARIMA and LSTM RNN // ITM Web of Conferences. – EDP Sciences, 2020. – Vol. 32. – p. 03017.
 6. Sikka D., Kumar C. N. S. V. Website Traffic Time Series Forecasting Using Regression Machine Learning // 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT). – IEEE, 2023. – pp. 246-250.
 7. Zhou K., et al. Comparative Study of Web Traffic Forecasting Time Series Based on a Statistical Model and a Generative Adversarial Model. – 2021. – Т. 213. – p. 106467.
 8. Matlakunta R. R. Web traffic prediction for online advertising : дис. – Auckland University of Technology, 2011.
 9. Telo J. Web Traffic Prediction Using Autoregressive, LSTM, and XGBoost Time Series Models // Applied Research in Artificial Intelligence and Cloud Computing. – 2020. – Т. 3. – №. 1. – pp. 1-15.
 10. Yao S., Hu C., Sun M. Prediction of web traffic based on wavelet and neural network // 2006 6th World Congress on Intelligent Control and Automation. – IEEE, 2006. – Т. 1. – pp. 4026-4028.
 11. Jun S. P., Sung T. E., Park H. W. Forecasting by analogy using the web search traffic // Technological Forecasting and Social Change. – 2017. – Т. 115. – pp. 37-51.
-