

Разработка алгоритма и программная реализация модуля корреляционно-регрессионного анализа и визуализации активности пользователей

И.В. Сребнова, С.Г. Симагина

Поволжский государственный университет телекоммуникаций и информатики, Самара

Аннотация: Цель статьи – программная реализация модуля для проведения анализа активности пользователей сайта на основе тепловой карты кликов, совместимого с отечественными веб-сервисами, объединяющего функционал корреляционно-регрессионного анализа и визуализацию в виде дашбордов до и после внесения изменений в элементы сайта. Весь функционал осуществляется непосредственно в сервисе веб-аналитики. На основании полученных данных по анализируемому элементу сайта принимается решение о корректировке дизайна и/или контента для повышения скорости кликов. Таким образом, предложенное решение позволяет расширить функционал сервиса веб-аналитики, снизить трудозатраты. Программный модуль был успешно протестирован. В результате проведения анализа и внесения необходимых корректировок на сайт, повысилось количество и соответственно скорость кликов.

Ключевые слова: активность пользователя, корреляционно-регрессионный анализ, дашборд, программный модуль, линия тренда, коэффициент детерминации.

Введение

В современном мире IT-технологии все активнее внедряются во все сферы человеческой деятельности. Невозможно уже представить области, в которых так или иначе не задействованы информационные технологии [1]. Они, безусловно, облегчают решение многих задач, двигают прогресс и совершенствуют мир в целом. Миллионы пользователей посредством интернета ежедневно получают неоценимую пользу и помощь в решении своих задач, включая заказы по интернету [2, 3]. Использование различных сайтов на постоянной основе для указанных целей неизбежно ведет к тому, что они требуют усовершенствования, доработки и разработки для более удобного их использования.

В связи с активным развитием веб-индустрии, свою активность набирает и область веб-аналитики. Все больше данных требуют обработки и анализа, для каждой кампании ставятся более глобальные задачи в веб-индустрии, которые требуют тщательного анализа и нахождения

зависимостей между различными показателями [4, 5]. С развитием технологий появляется потребность в создании новых качественных программных продуктов, которые будут оптимизировать процесс работы, избавлять от рутинной работы, сокращая время на сбор данных и повтора одних и тех же действий из раза в раз. Особое внимание требуется разработке отечественных программных продуктов, сервисов и т.д., которые будут совместимы с отечественным ПО и иметь расширенный функционал.

В частности, на основе проведенного сравнительного анализа является востребованным следующий функционал: анализ связи скорости кликов с целевым профилем пользователя, визуализация активности пользователей в реальном режиме, сравнения их активности до и после внесения корректировок в элементы сайта и контента [6, 7].

Цель статьи – разработка алгоритма и программного модуля для анализа и визуализация данных активности пользователей сайтов на основе тепловой карты.

Методология исследования

В работе использовались следующие основные методы и инструменты: корреляционный анализ, регрессионный анализ, метод наименьших квадратов, методы математического моделирования и принятия решений, сбор и обработка данных SEO-аналитики, методы и инструменты Яндекс. Метрики [8–10].

Также были использованы методы систематизации, обобщения и анализа веб-сервисов аналитики и научной литературы по тематике корреляционно-регрессионного анализа.

Результаты и их обсуждение

Проведенный анализ сервисов веб-аналитики показал, что актуальной становится задача разработки модуля для проведения анализа активности пользователей сайта на основе тепловой карты кликов, совместимого с наиболее распространенным отечественным веб-сервисом, например, Яндекс. Метрикой, объединяющий функционал корреляционно-регрессионного анализа и визуализацию в виде дашбордов непосредственно в сервисе веб - аналитики [11, 12]. Кроме этого, должна быть решена задача возможности сохранения и визуализации информации в виде дашбордов до и после внесения изменений в элементы сайта, по данным, полученным в результате анализа. Анализ проводится по числу кликов на элемент (N_{click}), полученных из Яндекс. Метрики. Далее производится расчет коэффициента кликабельности CTR (скорости кликов), который выражается в процентах [13]. Для элементов веб-сайта расчет делается до и после внесения изменений в элемент и определяется по формуле:

$$CTR = \frac{N_{click}}{N_{visit}} \cdot 100\%, \quad (1)$$

где N_{click} – число кликов на элемент; N_{visit} – количество визитов на странице.

Соответственно был разработан расширенный алгоритм для анализа активности пользователей, который на основе данных тепловой карты Яндекс. Метрики, реализует описанный ниже функционал. В первую очередь, выгружаются данные с API-Яндекс.Метрики, далее они сортируются и сегментируются по возрасту, полу, региону и устройству. Затем сегментированные данные выводятся на страницу программного модуля, по ним пользователь указывает параметры анализируемой аудитории, введенные данные сохраняются, и рассчитывается количество кликов с помощью программных библиотек. После чего определяется скорость кликов по формуле (1) и коэффициент достоверности аппроксимации (детерминации) для выбранной зависимости [9, 10]. Далее

данные выводятся на диаграмму рассеяния, рассчитывается уравнение линейной регрессии и строится линия тренда.

Если коэффициент достоверности аппроксимации меньше значения 0,5, то считается, что взаимосвязь между выбранными показателями отсутствует [10]. В этом случае необходимо провести анализ исследуемых элементов на странице сайта (например, изменить дизайн, расположение элемента и т.д.) и внести соответствующие корректировки для улучшения показателей скорости кликов. Для удобства проведения анализа, в программном модуле реализована возможность сохранения графика в архив. По прошествии времени, после внесения необходимых корректировок, возможно перейти на страницу архива, в которой сохранен изначальный результат проведения корреляционно-регрессионного анализа с сохранением построенного графика и расчетов. Для просмотра динамики данных после корректировок, необходимо на странице архива повторно провести анализ, после чего построится диаграмма рассеяния на основе новых показателей по тем же данным профиля пользователя.

Если коэффициент достоверности аппроксимации является больше или равно значения 0,5, в этом случае взаимосвязь между параметрами присутствует. При этом сохранение графика в архив возможно, но не является обязательным.

Проводить повторный анализ возможно до того момента, пока коэффициент достоверности аппроксимации будет больше или равен значению 0,5, или до момента принятия решения об удалении исследуемого элемента.

Алгоритм действий при разных показателях коэффициента достоверности аппроксимации показан на рис. 1.

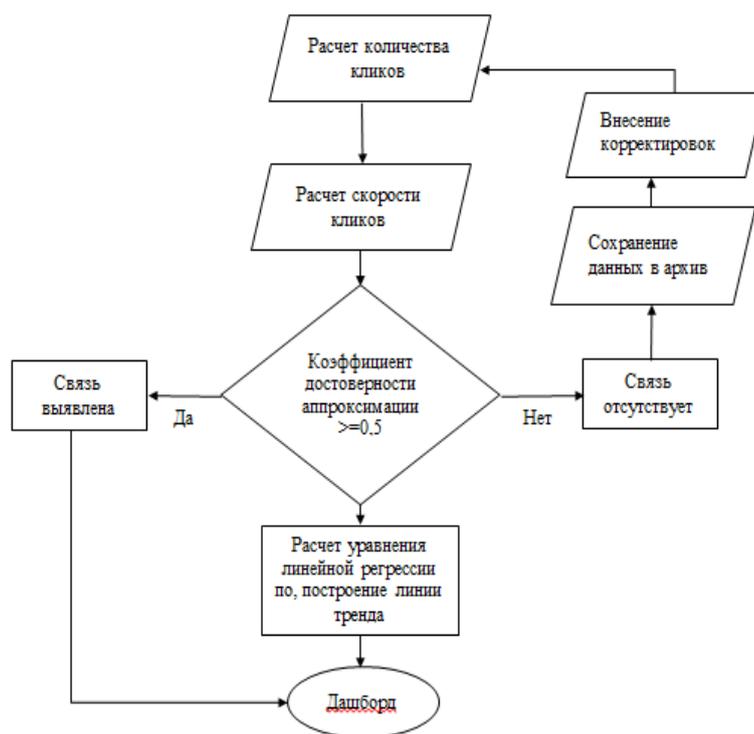


Рис. 1. – Алгоритм действий в зависимости от значения коэффициента достоверности аппроксимации

Задачей разработки программного модуля является сокращение ввода анализируемых данных на сторонних программах (сервисах, приложениях и т.д.), за счет чего уменьшаются сроки проведения анализа, расчет анализа в режиме реального времени, также мгновенного построения диаграммы рассеяния с линией тренда, расчетом коэффициента достоверности аппроксимации и линейного уравнения.

Для разработки программного модуля были использованы следующие технологии: языки программирования Python и JavaScript, среда разработки PyCharm, язык гипертекстовой разметки текста HTML, каскадные таблицы стилей CSS.

Программный модуль разработан в формате дашборда. Дашборд – формат графического интерфейса в виде аналитической панели. На дашборде

отображается несколько метрик, графиков, показателей процесса и динамики изменения данных.

Интерфейс программного модуля с расчетом кликов показан на рис. 2.

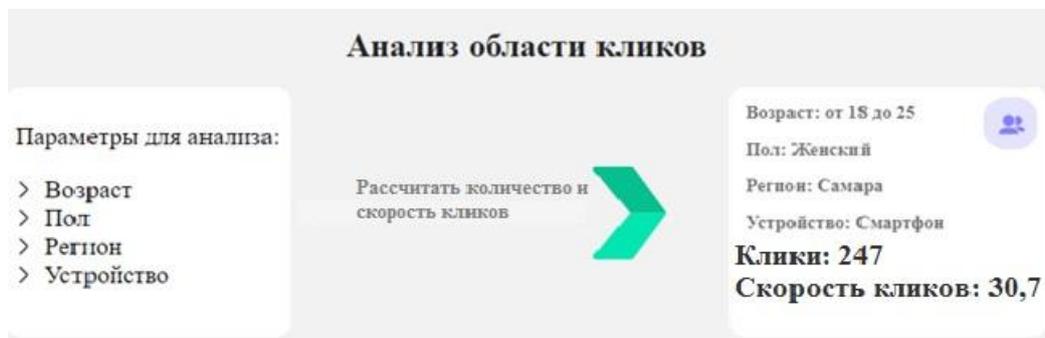


Рис 2. – Интерфейс программного модуля

Данные API, предоставленные с сервиса Яндекс. Метрика на основе тепловой карты и показателей вебвизора, обрабатываются и загружаются в разработанный программный модуль. Обработанные данные могут быть выведены на главную на страницу модуля (рис. 2).

На странице программного модуля представлены несколько блоков с параметрами, кнопки для проведения анализа, блок для вывода графика с данными и кнопки для работы с архивом. Для того, чтобы провести корреляционно-регрессионный анализ с помощью программного модуля, необходимо заполнить поля по анализируемой аудитории (возраст, пол, регион, устройство). Все поля обязательны для заполнения, таким образом, рассчитываются более точные показатели аудитории, а также снижается нагрузка при загрузке данных на сервере, что значительно ускоряет процесс проведения анализа. С помощью кнопки «рассчитать количество и скорость кликов» данные сохраняются и переносятся на другой блок с расчетом количества и скорости кликов по данным показателям. При этом учитывается расчет репрезентативности выборки. В случае, если поле не заполнено, появится окно уведомления с текстом «Для проведения анализа все поля должны быть заполнены!».

Если результат проведения анализа не устраивает и необходимо провести изменения на сайте для улучшения показателей, например, изменить дизайн или текст контента. В этом случае первоначальный график можно сохранить в архив, используя кнопки «Сохранить в архив» и «Перейти в архив» (рис. 3). Данные могут меняться в режиме реального времени.

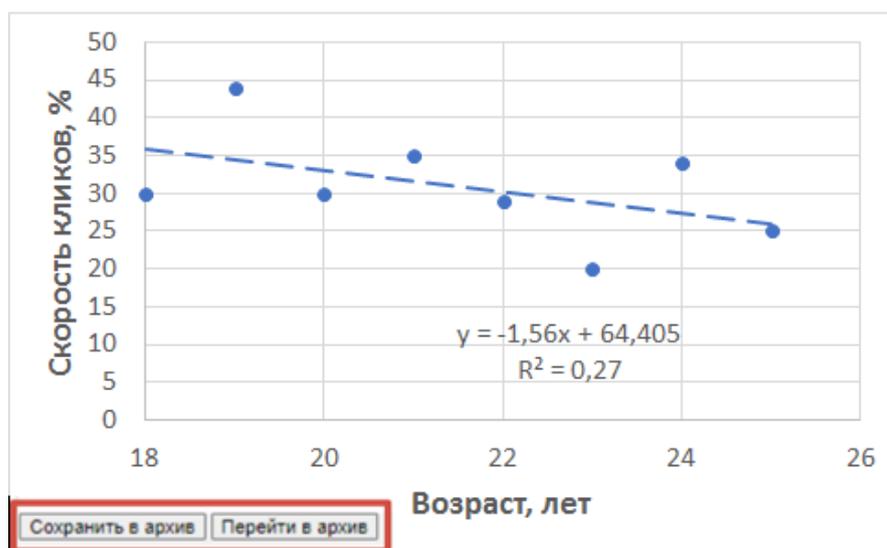


Рис. 3. – Результат проведения анализа с линией тренда

Было проведено тестирование разработанного модуля.

Для выявления взаимосвязи между показателями скорости кликов по рекламному баннеру и пользователями сайта, в программном модуле устанавливались параметры для проведения анализа, которые характеризуют профиль пользователя.

В нашем случае были заданы: возраст от 18 до 25, пол – женский, устройство - смартфон. Баннер был размещен на сайте с общим количеством пользователей 115867 человек. Предварительный расчет репрезентативности выборки осуществлялся при уровне надежности 95% и допустимой погрешности 7% и составил 196 пользователей. Уровень допустимой погрешности был выбран в рекомендуемом интервале для маркетинговых исследований 5-10%, так как предусматривалась ситуация, в которой

количество активных пользователей на начальной стадии может быть не достаточным [14, 15].

Во время проведения акции с использованием исследуемого баннера, общее количество кликов составило 247. После выгрузки и обработки данных анализировались показатели (скорость кликов) взаимодействия пользователей с рекламными баннерами в зависимости от возраста пользователя с заданным профилем.

При проведении анализа выяснилось, что коэффициент достоверности аппроксимации для полученной зависимости (рис. 3) составляет $R^2 = 0,27$, следовательно, взаимосвязь между выбранными показателями умеренная, линия тренда имеет отрицательный наклон. Такая активность пользователя не является удовлетворительной. Вследствие чего было принято решение изменить контент и расположение рекламного баннера.

После доработки рекламного баннера, была проведена повторная акция, анализ результатов которой выявил улучшения показателей по тем же параметрам анализируемой аудитории. В результате после внесенных корректировок в расположение рекламного баннера и наполненность контента количество кликов выросло до 440, средняя скорость кликов составила 47%. При этом для новой зависимости коэффициент достоверности аппроксимации составил $R^2 = 0,62$. Это указывает на то, что взаимосвязь скорости кликов и возрастом пользователей с выбранным профилем имеет высокую тесноту и является функциональной. Следовательно, внесение корректировок улучшило показатели активности пользователей.

Заключение

Был разработан алгоритм и осуществлена программная реализация модуля для проведения корреляционно-регрессионного анализа на основе данных тепловой карты сервиса Яндекс. Метрика. Разработанный

программный продукт позволяет проводить корреляционно-регрессионного анализ в режиме реального времени с целью снижения трудовых затрат на проведения анализа, выгрузку, обработку данных и принятия решения. Реализована возможность сохранения полученного результата в архив и проведения нового анализа на основе данных, после внесения необходимых изменений на сайте. Совместимость программного модуля с отечественным программным обеспечением позволяет расширить функционал сервиса веб-аналитики. Программный модуль был успешно протестирован. В результате проведения анализа и внесения необходимых корректировок повысилась скорость кликов.

Литература

1. Kotler P., Kartajaya H., Setiawan I. Marketing 4.0. Moving from Traditional to Digital. Hoboken: Wiley & Sons, 2017. 207 p.
 2. Сербиновский Б.Ю., Сербиновская А.А., Белоус М.А. Развитие коммуникаций и сбыта продукции и услуг ресторанной сети с использованием IT-технологий (опыт моделирования и проектирования многофункционального веб-сайта). Часть 1 // Инженерный вестник Дона, 2013, № 2. URL: ivdon.ru/ru/magazine/archive/n2y2013/1610.
 3. Симагина С. Г., Евстафьева В.А. Исследование развития информационного общества в сфере электронной коммерции // Отходы и ресурсы, 2023, Т. 10, № 2. URL: resources.today/PDF/07ECOR223.pdf.
 4. Мальцева Ю.А., Котляревская И.В., Багинская В.А. Коммуникационные инструменты маркетинга в интернет пространстве. Екатеринбург: Изд-во Уральского университета, 2021. 73 с.
 5. Фейгель М.Л. Теоретические аспекты ситуативно-прогнозной ориентации маркетинговых коммуникаций в комплексе маркетинга компетенции при продвижении управленческих сервисов // Инженерный вестник Дона, 2013, №4. URL: ivdon.ru/magazine/archive/n4y2013/1882.
-

6. Сервис Яндекс.Метрика. URL: yandex.ru/support/metrika/ (дата обращения: 22.08.2024)
7. Шишакова Ю. В. , Башкина Н.А. Современные факторы эффективности применения цифровых технологий в продвижении брендов // E-Management. 2022. Т. 5. № 4. С. 106-112.
8. Щепакин М. Б., Соболев А.В. К развитию рекламно-маркетингового инструментария в среде // Экономика, предпринимательство и право, 2021, Т. 11, № 8. С. 2005-2026. URL: DOI 10.18334/epp.11.8.113119.
9. Эконометрика/ Под ред. И.И. Елисеевой. М.: Финансы и статистика, 2006. 576 с.
10. Zacks S. The Theory of Statistical Inference. Hoboken: Wiley & Sons Inc, 1971. 626 p.
11. Данилов Н.А. Модели и алгоритмы анализа и визуализации данных активности пользователей информационных систем: автореф. дис.... канд. тех. наук. Саратов, 2019. 20 с.
12. Симагина С. Г., Сребнова И.В. Сравнительный анализ сервисов веб-аналитики и разработка алгоритма их расширенного функционирования на основе отечественных решений // Вестник Самарского государственного экономического университета. 2024. № 7(237). С. 44-50.
13. Санталова М.С. Инновационное управление персоналом: цифровые технологии и развитие креативности. М.: Изд-во «Дашков и К⁰», 2022. 208 с.
14. Маркетинг / Под ред. Н.Д. Эриашвили, Ю.А. Цыпкина. М.: ЮНИТИ-ДАНА, 2023. 184с.
15. Рольбина Е.С. Маркетинговые исследования, сегментация, позиционирование. Казань: Изд-во КГФЭИ, 2011. 244 с.

References

1. Kotler P., Kartajaya H., Setiawan I. Marketing 4.0. Moving from Traditional to Digital. Hoboken: Wiley & Sons , 2017. 207 p.



2. Serbinovskij B.Yu., Serbinovskaya A.A., Belous M.A. Inzhenernyj vestnik Dona, 2013, № 2. URL: ivdon.ru/ru/magazine/archive/n2y2013/1610.
3. Simagina S. G., Evstaf'eva V.A. Othody i resursy, 2023, T. 10, № 2. URL: resources.today/PDF/07ECOR223.pdf.
4. Mal'ceva Yu.A., Kotlyarevskaya I.V., Baginskaya V.A. Kommunikacionnye instrumenty marketinga v internet prostranstve [Communication tools for marketing in the Internet space]. Ekaterinburg: Izd-vo Ural'skogo universiteta, 2021. 73 p.
5. Fejgel' M.L. Inzhenernyj vestnik Dona, 2013, №4. URL: ivdon.ru/magazine/archive/n4y2013/1882.
6. Servis Yandeks.Metrika [Yandex.Metrica service]. URL: yandex.ru/support/metrica/ (date assessed: 22.08.2024)
7. Shishakova Yu. V., Bashkina N.A. E-Management. 2022. T. 5. № 4. P. 106-112.
8. Shchepakina M. B., Sobol' A.V. Ekonomika, predprinimatel'stvo i pravo, 2021, T. 11, № 8. pp. 2005-2026. URL: DOI 10.18334/epp.11.8.113119.
9. Ekonometrika [Econometrics]. Pod red. I.I. Eliseevoj. M.: Finansy i statistika, 2006. 576 p.
10. Zacks S. The Theory of Statistical Inference. Hoboken: Wiley & Sons Inc, 1971. 626 p.
11. Danilov N.A. Modeli i algoritmy analiza i vizualizacii dannyh aktivnosti pol'zovatelej informacionnyh sistem [Models and algorithms for analysis and visualization of information system user activity data]: avtoref. dis.... kand. tekhn. nauk. Saratov, 2019. 20 p.
12. Simagina S. G., Srebnova I.V. Vestnik Samarskogo gosudarstvennogo ekonomicheskogo universiteta. 2024. № 7(237). pp. 44-50.
13. Santalova M.S. Innovacionnoe upravlenie personalom: cifrovye tekhnologii i razvitiye kreativnosti [Innovative HR Management: Digital



Technologies and Creativity Development]. М.: Izd-vo «Dashkov i K⁰», 2022. 208 p.

14. Marketing [Marketing]. Pod red. N.D. Eriashvili, Yu.A. Cypkina. М.: YuNITI-DANA, 2023. 184 p.

15. Rol'bina E.S. Marketingovye issledovaniya, segmentaciya, pozicionirovanie [Marketing research, segmentation, positioning]. Kazan': Izd-vo KGFEI, 2011. 244 p.

Дата поступления: 9.09.2024

Дата публикации: 20.10.2024