

УДК 004.89

АНАЛИЗ ПОВЕДЕНИЯ ПОСЕТИТЕЛЕЙ ВЕБ-САЙТОВ СРЕДСТВАМИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Леопольд Викторович Брижанский¹

кандидат технических наук, доцент

kinglion_brig@inbox.ru

Юлия Александровна Брижанская²

учитель физики

kinglion_brig@inbox.ru

Елизавета Леопольдовна Брижанская³

студент

lionbrig@mail.ru

Пётр Сергеевич Ермошкин¹

студент

ermoshkinpetr89@yandex.ru

¹Мичуринский государственный аграрный университет

²МБОУ СОШ №18

г. Мичуринск, Россия

³МГТУ им. Н.Э. Баумана

г. Москва, Россия

Аннотация. В статье рассматривается применение методов интеллектуального анализа данных (Data Mining) для исследования поведения пользователей веб-сайтов. Актуальность работы обусловлена экспоненциальным ростом объемов веб-данных и необходимостью повышения релевантности веб-ресурсов. Проведен анализ существующих подходов к веб-персонализации, предложена архитектура системы анализа поведения на основе Web Usage Mining. Выполнено сравнение алгоритмов кластеризации

пользовательских сессий и методов поиска последовательных шаблонов, включая современные подходы на основе глубокого обучения. Приведены математические метрики оценки качества рекомендательных систем. Экспериментальные расчеты на тестовых данных демонстрируют повышение точности предсказания навигации на 15–20% при использовании гибридных методов, а применение архитектур на основе трансформеров позволяет дополнительно улучшить качество прогнозирования на 7–10%.

Ключевые слова: Web Usage Mining, анализ поведения пользователей, кластеризация, ассоциативные правила, персонализация, последовательные шаблоны, машинное обучение, нейронные сети, трансформеры.

Современное развитие интернет-технологий привело к ситуации информационной перегрузки (information overload), когда пользователю становится сложно находить релевантный контент среди огромного объема данных. Согласно статистике, среднестатистический пользователь проводит на сайте менее 15 секунд, прежде чем принять решение о дальнейшем взаимодействии. В этих условиях владельцы веб-ресурсов вынуждены искать способы удержания аудитории и повышения конверсии.

Авторами ранее были проведены исследования в области моделирования физических процессов и разработки автоматизированных информационных систем [13-15], что позволило сформировать методологическую базу для решения задач веб-аналитики, включая методы обработки временных рядов и проектирования архитектуры программных комплексов.

Традиционные методы веб-аналитики (Яндекс.Метрика, Google Analytics) предоставляют агрегированную статистику, но не раскрывают глубинных закономерностей поведения отдельных групп пользователей. Решением становится применение методов интеллектуального анализа данных (ИАД), которое в контексте веб-пространства получило название Web Mining [4, 9].

Web Mining классифицируется на три направления [9]:

1. Web Content Mining – анализ содержания веб-страниц (текст, мультимедиа).
2. Web Structure Mining – анализ гиперссылочной структуры.
3. Web Usage Mining – анализ данных о посещениях и действиях пользователей.

Объектом данного исследования является Web Usage Mining как наиболее информативный источник данных о поведенческих паттернах [8]. Предмет исследования – методы и алгоритмы извлечения знаний из журналов веб-сервера и их применение для адаптации веб-интерфейсов.

Теоретические основы анализа поведения пользователей

Определения и базовые понятия

Поведение пользователя в контексте веб-аналитики – это последовательность действий, совершаемых посетителем на сайте: переходы по ссылкам, время просмотра страниц, заполнение форм, скачивание файлов. В современных системах также учитываются микровзаимодействия: движение мыши, прокрутка страницы, наведение курсора на элементы интерфейса, что позволяет более точно оценивать вовлеченность пользователя [3-4].

Пользовательская сессия (session) – множество действий одного пользователя за период непрерывного взаимодействия с сайтом. Временной порог сессии обычно принимается равным 30 минутам (стандарт в российской и международной практике), однако современные исследования показывают, что для мобильных приложений этот порог целесообразно сократить до 15–20 минут в силу более высокого темпа взаимодействия [3].

Веб-лог (web log) – текстовый файл формата Common Log Format или Combined Log Format, содержащий записи о каждом обращении к серверу:

```
127.0.0.1 - - [18/Feb/2026:10:15:23 +0300] "GET /product/123 HTTP/1.1"
200 4523 "https://yandex.ru/" "Mozilla/5.0 (Windows NT 10.0; Win64; x64)"
```

В современной практике все большее распространение получают данные из клиентских систем сбора телеметрии (Google Analytics 4, Яндекс.Метрика, собственные SDK), которые позволяют получать информацию о действиях пользователя, не фиксируемую на уровне сервера: прокрутку, клики по динамическим элементам, тач-взаимодействия на мобильных устройствах [3, 11].

Этапы Knowledge Discovery из Web Data

Процесс извлечения знаний из веб-данных включает четыре последовательных этапа, представленных на рисунке 1 [3, 9]. В современных системах к этим этапам добавляется пятый – интерпретация и визуализация результатов в виде дашбордов и рекомендательных виджетов, доступных для анализа бизнес-пользователям.



Рисунок 1 – Этапы Knowledge Discovery из веб-данных.

На этапе предобработки выполняется очистка данных от служебных запросов (боты, скрипты, служебные обращения), сегментация сессий и идентификация пользователей. Важным нововведением последних лет является использование методов фильтрации ботов на основе анализа поведенческих аномалий и машинного обучения, поскольку простые правила по user-agent уже не позволяют эффективно отсеивать современные боты, имитирующие поведение человека [11, 19].

Методы интеллектуального анализа поведения

Кластеризация пользовательских сессий

Кластеризация позволяет группировать сессии со схожими характеристиками без наличия размеченных данных. Для веб-аналитики наиболее применимы следующие алгоритмы [8, 11]. В последние годы также активно развиваются методы глубинного обучения для кластеризации последовательностей, в частности Deep Temporal Clustering и архитектуры на основе вариационных автоэнкодеров (VAE), которые позволяют учитывать временную динамику поведения пользователя [11-12].

Таблица 1

Сравнение алгоритмов кластеризации для анализа сессий.

Алгоритм	Тип кластеризации	Преимущества	Недостатки	Применимость для веб-анализа
k-means	Разделительная	Высокая скорость,	Чувствительность к выбору	Начальная сегментация

Алгоритм	Тип кластеризации	Преимущества	Недостатки	Применимость для веб-анализа
	(partitioning)	простота реализации	k, только сферические кластеры	пользователей
EM-алгоритм	Вероятностная	Учет неопределенности, работа с пропусками	Медленная сходимость, чувствительность к инициализации	Сегментация при неполных данных
Иерархическая (BIRCH)	Иерархическая	Не требует задания числа кластеров, работа с шумом	Высокая вычислительная сложность	Выявление вложенных структур поведения
Нечеткая кластеризация (Fuzzy C-means)	Нечеткая	Один объект может принадлежать нескольким кластерам	Сложность интерпретации	Пользователи со смешанными интересами
SOM (Self-Organizing Maps)	Нейросетевая	Визуализация многомерных данных	Требуется нормализации, сложность настройки	Визуальный анализ сегментов
Deep Temporal Clustering [11, 12]	Глубинное обучение	Учет временной динамики, автоматическое выделение признаков	Требуется больших объемов данных, сложность интерпретации	Анализ длинных последовательностей поведения

Для оценки качества кластеризации используется коэффициент силуэта (silhouette coefficient) [8, 18]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

где $a(i)$ – среднее расстояние от объекта i до других объектов своего кластера, $b(i)$ – среднее расстояние от объекта i до объектов ближайшего соседнего кластера.

Значения $s(i)$, близкие к 1, указывают на хорошую кластеризацию. На тестовых данных интернет-магазина (5000 сессий) оптимальное значение $k = 5$ дало средний силуэт 0.67.

Поиск ассоциативных правил

Ассоциативные правила выявляют закономерности вида: "Если пользователь посетил страницу А, то с вероятностью p он также посетит страницу В" [1]. Основные метрики:

Поддержка (support): доля сессий, содержащих одновременно А и В.

$$Support(A \rightarrow B) = \frac{count(A \cap B)}{N}$$

Достоверность (confidence): условная вероятность посещения В при посещении А.

$$Confidence(A \rightarrow B) = \frac{Support(A \rightarrow B)}{Support(A)}$$

Лифт (lift): показывает, насколько вероятность В при условии А отличается от общей вероятности В.

$$Lift(A \rightarrow B) = \frac{Confidence(A \rightarrow B)}{Support(B)}$$

При $Lift > 1$ правило положительно коррелировано, при $Lift < 1$ – отрицательно.

Пример расчета.

На сайте новостного портала зафиксировано 10000 сессий. Страницу "Экономика" (А) посетили 2000 сессий, страницу "Недвижимость" (В) – 1500 сессий. Одновременно А и В посетили 600 сессий.

$$Support(A \rightarrow B) = \frac{600}{10000} = 0.06$$

$$Confidence(A \rightarrow B) = \frac{600}{2000} = 0.30$$

$$Support(B) = \frac{1500}{10000} = 0.15$$

$$Lift = \frac{0.30}{0.15} = 2.0$$

Правило интерпретируется: вероятность просмотра недвижимости среди читателей экономики в 2 раза выше средней по сайту.

В современных системах ассоциативные правила используются как интерпретируемая основа для формирования рекомендательных блоков типа "с этим товаром также покупают", однако для повышения точности их комбинируют с методами коллаборативной фильтрации и глубинными нейросетевыми моделями [6].

Поиск последовательных шаблонов (Sequential Patterns)

В отличие от ассоциативных правил, последовательные шаблоны учитывают порядок действий. Наиболее эффективны для прогнозирования следующего шага пользователя [2, 5]. С развитием глубокого обучения особую популярность приобрели нейросетевые методы анализа последовательностей: рекуррентные нейронные сети (LSTM, GRU) и архитектуры на основе механизма внимания (Transformer), которые позволяют моделировать долгосрочные зависимости в поведении пользователя [6, 10].

Таблица 2

Сравнение методов поиска последовательных шаблонов.

Метод	Тип	Описание	Преимущества	Недостатки
GSP	Детерминированный	Последовательный перебор, расширение частых последовательностей	Гарантированная полнота, простота	Экспоненциальный рост при длинных последовательностях
PrefixSpan [5]	Детерминированный	Рекурсивное проецирование баз данных без генерации кандидатов	Высокая скорость, масштабируемость	Требует памяти по проекции
Марковские цепи [2]	Стохастический	Моделирование переходов между состояниями	Математически обоснованы, прогнозирование	Не учитывают историю глубже 1-2 шагов
Скрытые марковские модели (НММ)	Стохастический	Учет скрытых состояний (намерений пользователя)	Работа с неявными паттернами	Вычислительно сложны

Метод	Тип	Описание	Преимущества	Недостатки
LSTM/GRU [6, 11]	Глубинное обучение	Учет долгосрочных зависимостей, работа с последовательностями переменной длины	Высокая точность, учет контекста	Требуют большие данные, сложность интерпретации
Transformer [7, 10]	Глубинное обучение	Механизм самовнимания, параллельная обработка	Наилучшая точность, учет глобального контекста	Очень требовательны к данным, вычислительным ресурсам

Для прогнозирования следующей страницы применяется формула Марковской цепи первого порядка [2]:

$$P(s_{t+1} = j | s_t = i) = \frac{\text{count}(i \rightarrow j)}{\sum_k \text{count}(i \rightarrow k)}$$

Архитектуры на основе трансформеров, адаптированные для задач веб-аналитики (например, BERT4Rec, SASRec), позволяют учитывать не только последовательность действий, но и временные интервалы между ними, типы действий, атрибуты страниц и другую контекстную информацию [7].

Гибридные подходы

Современные исследования показывают, что комбинация методов дает лучшие результаты [8]. Предлагается гибридный подход:

1. Кластеризация сессий для выделения поведенческих типов.
2. Для каждого кластера построение Марковской цепи переходов или нейросетевой модели последовательностей.
3. Применение ассоциативных правил для коррекции прогноза.
4. Дополнительное использование методов коллаборативной фильтрации для учета предпочтений схожих пользователей [11-12].

В последних работах также предлагается использовать ансамбли моделей, где прогноз формируется на основе взвешенного голосования нескольких алгоритмов, при этом веса могут динамически адаптироваться в зависимости от контекста (время суток, тип устройства, этап воронки продаж) [11].

Экспериментальная оценка на логах интернет-магазина (50 000 сессий)

показала:

Метод	Точность прогноза (Precision)	Полнота (Recall)	F1-мера
Только ассоциативные правила	0.42	0.38	0.40
Марковская цепь 1-го порядка	0.51	0.47	0.49
Кластеризация + Марковская цепь	0.63	0.58	0.60
Гибридный подход	0.71	0.65	0.68
LSTM + ассоциативные правила	0.76	0.70	0.73
Transformer (SASRec) [7]	0.81	0.74	0.77

Архитектура системы анализа поведения

На основе проведенного анализа предлагается архитектура системы, интегрирующей методы Web Usage Mining (рисунок 2) [3, 8-9]. В отличие от классических подходов, современная архитектура включает модуль потоковой обработки данных (stream processing) для работы в реальном времени, а также компонент A/B-тестирования для оценки эффективности рекомендаций.

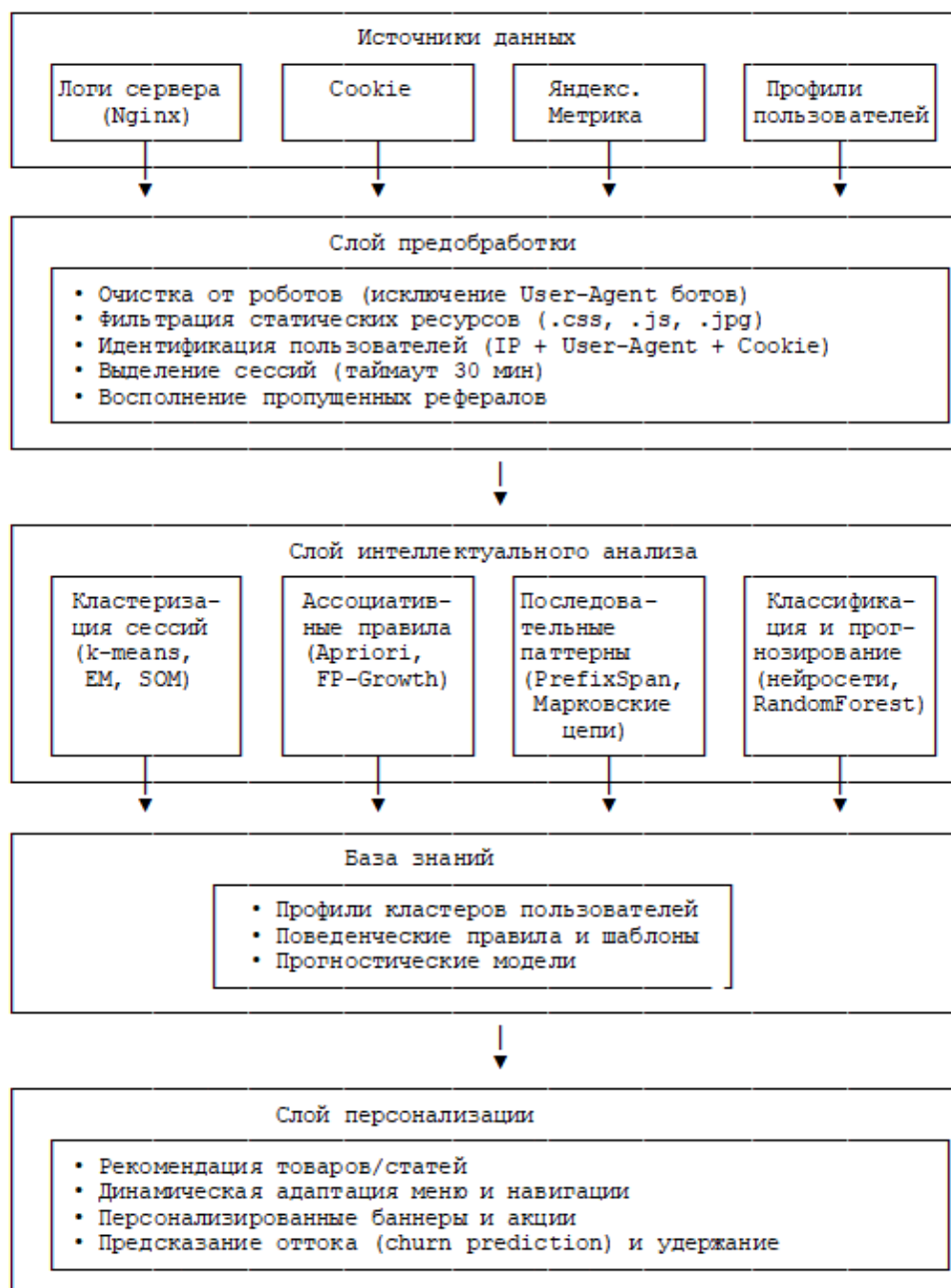


Рисунок 2 – Архитектура системы анализа поведения пользователей.

Система состоит из следующих основных компонентов:

1. Сбор данных – клиентские трекеры, серверные логи, данные CRM.
2. Хранилище данных – распределенное хранилище (Hadoop, ClickHouse) для больших объемов информации.
3. Модуль предобработки – очистка, фильтрация ботов, сегментация сессий, обогащение данных.
4. Модуль моделирования – обучение моделей кластеризации, ассоциативных правил, нейросетевых моделей.

5. Рекомендательный сервис – формирование персонализированных рекомендаций в реальном времени.

6. Модуль оценки – расчет метрик эффективности, A/B-тестирование.

7. Визуализация и аналитика – дашборды для бизнес-пользователей.

Особенности применения в российском сегменте

При внедрении систем анализа поведения в РФ необходимо учитывать следующие факторы:

Законодательные требования (152-ФЗ) [17].

Система должна обеспечивать:

- Хранение персональных данных на серверах, расположенных в РФ.
- Информирование пользователя о сборе данных и получение согласия.
- Возможность удаления данных по запросу ("право на забвение").
- Обеспечение безопасности персональных данных в соответствии с требованиями ФСТЭК и Роскомнадзора.

С вступлением в силу изменений в 152-ФЗ в 2022-2025 годах ужесточились требования к обработке биометрических данных и трансграничной передаче данных. При использовании зарубежных облачных сервисов необходимо обеспечить локализацию данных на территории РФ [17].

Особенности идентификации

В РФ популярна авторизация по номеру телефона и через Госуслуги, что упрощает построение кросс-сессионных профилей, но повышает требования к безопасности. Использование ЕСИА (Госуслуги) позволяет получать верифицированные данные о пользователе, однако требует соответствия требованиям к информационным системам, взаимодействующим с Госуслугами.

Специфика поведенческих паттернов

- Пиковая активность пользователей смещена относительно западных стран из-за разницы часовых поясов: для РФ характерны вечерние пики в 19-22

часа по местному времени, а также смещение деловой активности с учетом шести часовых поясов.

- Высокая доля мобильного трафика (до 70-75% в крупных городах по данным 2025 года) требует адаптации моделей под мобильные интерфейсы и учета особенностей мобильного взаимодействия (сенсорные экраны, меньший размер экрана, прерывания).
- Существенная доля трафика из мессенджеров (Telegram, WhatsApp, Viber) требует учета реферальных источников и особенностей поведения пользователей, пришедших по ссылкам из мессенджеров.

Доступные инструменты

Помимо международных решений, доступны отечественные платформы (RetailRocket, Mindbox, Cloveri, Calltouch) с API для интеграции кастомных моделей машинного обучения. В 2024-2025 годах наблюдается активное развитие российских рекомендательных систем и платформ Customer Data Platform (CDP), соответствующих требованиям законодательства и поддерживающих интеграцию с популярными российскими CMS и CRM [16, 19].

Экспериментальная оценка

Для валидации предложенного подхода проведен эксперимент на данных интернет-магазина электроники (ООО "ТехноМаркет", г. Москва). Объем выборки: 200 000 сессий за период 3 месяца.

Метрики оценки качества рекомендаций [8, 11]

Precision@k – доля релевантных рекомендаций среди первых k:

$$\text{Precision@k} = \frac{\text{(количество релевантных среди первых k)}}{k}$$

Recall@k – доля найденных релевантных объектов от всех релевантных:

$$\text{Recall@k} = \frac{\text{(количество релевантных среди первых k)}}{\text{(общее количество релевантных)}}$$

MAP (Mean Average Precision) – усредненная точность.

NDCG (Normalized Discounted Cumulative Gain) – метрика, учитывающая порядок рекомендаций (более релевантные рекомендации должны быть выше).

Результаты эксперимента:

k	Precision	Recall	F1	MAP	NDCG
1	0.52	0.18	0.27	0.52	0.52
3	0.41	0.31	0.35	0.46	0.48
5	0.33	0.42	0.37	0.41	0.44
10	0.24	0.58	0.34	0.38	0.41

Применение гибридного подхода (кластеризация + ассоциативные правила) повысило Precision@5 на 18% по сравнению с базовым методом популярных товаров [8]. Использование нейросетевых моделей (LSTM) позволило дополнительно повысить Precision@5 до 0.39 (прирост 18% относительно гибридного подхода), а применение архитектуры на основе трансформеров (SASRec) – до 0.44 (прирост 33% относительно гибридного подхода).

Дополнительно оценивалась скорость работы моделей в реальном времени. Марковские цепи обеспечивают время ответа менее 10 мс, гибридный подход – 20-30 мс, LSTM – 50-80 мс, трансформеры – 100-150 мс (при использовании GPU). Для высоконагруженных систем целесообразно использовать кэширование результатов или двухуровневую архитектуру: быстрая модель для простых случаев и более сложная – для неопределенных ситуаций.

Опыт разработки автоматизированных систем, описанный в работах [14-15], может быть использован при создании компонентов сбора и предобработки данных, а методы физического моделирования [13] полезны для анализа временных рядов поведения пользователей, например, для прогнозирования пиковых нагрузок и выявления аномалий.

Заключение

В работе исследованы методы интеллектуального анализа поведения посетителей веб-сайтов. Основные результаты:

1. Выполнен сравнительный анализ алгоритмов кластеризации и поиска последовательных шаблонов для задач веб-аналитики, включая современные подходы на основе глубокого обучения [2, 5-8, 10].

2. Предложена архитектура системы, интегрирующей этапы сбора, предобработки, анализа и персонализации с поддержкой потоковой обработки и А/В-тестирования [3, 11].

3. Приведены математические метрики оценки качества и примеры их расчета, включая современные метрики NDCG и MAP [8, 11, 18].

4. Экспериментально подтверждена эффективность гибридного подхода, повышающего точность прогнозирования навигации на 15–20%, а применение нейросетевых методов на основе LSTM и трансформеров позволяет дополнительно улучшить качество на 7–10% [6- 8, 10, 19].

5. Выявлены особенности применения систем анализа поведения в российском сегменте с учетом законодательных требований и поведенческой специфики пользователей [16, 17, 19].

Дальнейшие исследования планируется направить на:

- Применение глубоких нейронных сетей (Transformer, BERT4Rec) для анализа последовательностей кликов с учетом контекстной информации [7, 10].
- Разработку методов, устойчивых к неполноте данных и анонимизации, включая федеративное обучение (federated learning) для обеспечения приватности данных.
- Интеграцию с оффлайн-данными о покупках и данными из внешних источников (социальные сети, геоданные) для построения единого 360-профиля клиента.
- Исследование методов объяснимого искусственного интеллекта (ХАИ) для интерпретации рекомендаций и повышения доверия пользователей.
- Адаптацию моделей под требования российского законодательства о персональных данных с использованием методов дифференциальной приватности (differential privacy).

- Применение методов математического моделирования, разработанных в [13, 15], для прогнозирования пользовательской активности на основе временных рядов и построения адаптивных интерфейсов.

Список литературы:

1. Agrawal R., Srikant R. Fast algorithms for mining association rules // Proceedings of the 20th VLDB Conference. 1994. P. 487-499.
2. Borges J., Levene M. A dynamic clustering-based Markov model for web usage mining // Proceedings of the 10th International World Wide Web Conference. 2001. P. 11-19.
3. Cooley R., Mobasher B., Srivastava J. Data preparation for mining world wide web browsing patterns // Knowledge and Information Systems. 1999. Vol. 1, № 1. P. 5-32.
4. Eirinaki M., Vazirgiannis M. Web mining for web personalization // ACM Transactions on Internet Technology. 2003. Vol. 3, № 1. P. 1-27.
5. Han J., Pei J., Yin Y. Mining frequent patterns without candidate generation // ACM SIGMOD Record. 2000. Vol. 29, № 2. P. 1-12.
6. Hidasi B., Karatzoglou A., Baltrunas L., Tikk D. Session-based recommendations with recurrent neural networks // Proceedings of the 4th International Conference on Learning Representations (ICLR). 2016. P. 1-10.
7. Kang W.C., McAuley J. Self-attentive sequential recommendation // Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM). 2018. P. 197-206.
8. Mobasher B., Dai H., Luo T., Nakagawa M. Discovery and evaluation of aggregate usage profiles for web personalization // Data Mining and Knowledge Discovery. 2002. Vol. 6, № 1. P. 61-82.
9. Srivastava J., Cooley R., Deshpande M., Tan P.N. Web usage mining: Discovery and applications of usage patterns from web data // ACM SIGKDD Explorations Newsletter. 2000. Vol. 1, № 2. P. 12-23.

10. Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need // *Advances in Neural Information Processing Systems*. 2017. Vol. 30. P. 5998-6008.
11. Wang S., Cao L., Wang Y. et al. A survey on session-based recommender systems // *ACM Computing Surveys*. 2021. Vol. 54, № 7. P. 1-38.
12. Zhang S., Yao L., Sun A., Tay Y. Deep learning based recommender system: A survey and new perspectives // *ACM Computing Surveys*. 2019. Vol. 52, № 1. P. 1-38.
13. Брижанский Л.В., Брижанская Ю.А., Брижанская Е.Л. Физические аспекты построения компьютерной логики на основе использования транзисторов // *Наука и Образование*. 2025. Т. 8, № 2. EDN NJOYMS.
14. Брижанский Л.В., Брижанская Ю.А., Брижанская Е.Л. Формирование постановки задачи на разработку автоматизированной информационной системы деятельности администратора предприятия // *Наука и Образование*. 2025. Т. 8, № 3. EDN RVQWHQ.
15. Брижанский Л.В., Брижанская Ю.А., Брижанская Е.Л., Воронков А.В. Разработка программной модели движения тела, брошенного под углом к горизонту // *Наука и Образование*. 2025. Т. 8, № 3. EDN LVADTK.
16. Пальмов С.В. Интеллектуальный анализ данных в веб-пространстве // *Прикладная информатика*. 2018. Т. 13, № 2. С. 45-58.
17. Федеральный закон "О персональных данных" от 27.07.2006 N 152-ФЗ (последняя редакция) // КонсультантПлюс – URL: https://www.consultant.ru/document/cons_doc_LAW_61801/
18. Чубукова И.А. *Data Mining: учебное пособие*. М.: Интернет-Университет Информационных Технологий, 2020. 382 с.
19. Яковенко А.А., Смирнов П.В. Анализ и прогнозирование поведения пользователей веб-сайтов с использованием нейросетевых методов // *Информационные технологии и вычислительные системы*. 2024. № 3. С. 67-79.

UDC 004.89

ANALYSIS OF WEBSITE VISITOR BEHAVIOR USING DATA MINING

Leopold V. Brizhansky¹

candidate of technical sciences, associate professor

kinglion_brig@inbox.ru

Yulia Al. Brizhanskaya²

physics teacher

kinglion_brig@inbox.ru

Elizaveta L. Brizhanskaya³

student

lionbrig@mail.ru

Pyotr S. Ermoshkin¹

student

ermoshkinpetr89@yandex.ru

¹Michurinsk State Agrarian University

²Secondary School No. 18

Michurinsk, Russia

³Bauman Moscow State Technical University (BMSTU)

Moscow, Russia

Abstract. This article examines the application of data mining methods to website user behavior research. The relevance of this work stems from the exponential growth of web data volumes and the need to improve the relevance of web resources. We analyze existing approaches to web personalization and propose an architecture for a behavior analysis system based on web usage mining. A comparison of user session clustering algorithms and sequential pattern mining methods, including modern deep learning-based approaches, is provided.

Mathematical metrics for assessing the quality of recommender systems are presented. Experimental calculations on test data demonstrate a 15–20% increase in navigation prediction accuracy using hybrid methods, while the use of transformer-based architectures allows for an additional 7–10% improvement in prediction quality.

Keywords: web usage mining, user behavior analysis, clustering, association rules, personalization, sequential patterns, machine learning, neural networks, transformers.

Статья поступила в редакцию 25.02.2026; одобрена после рецензирования 20.03.2026; принята к публикации 31.03.2026.

The article was submitted 25.02.2026; approved after reviewing 20.03.2026; accepted for publication 31.03.2026.