

УДК 330.47: 519.246.5

DOI 10.5281/zenodo.18048614

**УКРАИНСКИЙ Константин Викторович<sup>1</sup>**

<sup>1</sup>Школа вычислительных социальных наук Европейского университета в Санкт-Петербурге, ул. Гагаринская, д. 6/1, литера А, Санкт-Петербург, Россия, 191187

## **НЕЙРОННЫЕ ПРОЦЕССЫ ХОУКСА ДЛЯ ОЦЕНКИ ПАРАМЕТРОВ ИНТЕНСИВНОСТИ И ОПТИМИЗАЦИИ МАРКЕТИНГОВЫХ СТРАТЕГИЙ В УСЛОВИЯХ ОГРАНИЧЕННОГО ДОСТУПА К ДАННЫМ**

Статья посвящена разработке интерпретируемой нейронной модели на основе процессов Хоукса с механизмами внимания для анализа динамики прослушиваний музыкальных произведений на цифровых платформах. В работе обосновывается актуальность оценки эффективности маркетинговых стратегий в условиях ограниченного доступа к данным, вызванного регуляторными ограничениями. Рассмотрены текущие проблемы сбора и обработки информации, включая временные метки событий и категориальные переменные, такие как источник прослушивания, возрастная группа и время суток. Изучены эмпирические исследования пользовательского поведения, выделены группы «специалистов», «генералистов» и интенсивных пользователей, подчеркивающие необходимость учета разнообразия потребления и эффектов самовозбуждения.

Детально проанализирована эконометрическая модель повторяющегося дискретного выбора, представлена динамика спроса как агрегированного счетчика событий с учетом доли пересматривающих потребителей и эндогенных факторов. Охарактеризованы ограничения традиционных подходов, включая модели Pareto/NBD и классические процессы Хоукса, связанные с независимостью событий и экспоненциальным ростом параметров. Выделены преимущества предлагаемой модели, интегрирующей векторные представления и трансформерную архитектуру для захвата самовозбуждения и ингибирования.

Важным моментом исследования выступает методология, включающая предобработку данных, моделирование интенсивности с расширением на маркированные события, разделение на обучающую и валидационную выборки, оценку качества через остаточный анализ и метрики регрессии. Представлены механизмы раздельного обучения категорий событий и оценки важности независимых переменных через веса внимания. Статья содержит анализ ядер воздействия, суммарного влияния типов событий и оптимальных временных лагов. Планируемые результаты отражены в таблицах и диаграммах, демонстрирующих асимметрию влияний и практическую ценность для оптимизации кампаний.

Сделаны выводы о способности модели преодолевать ограничения данных, обеспечивая точное прогнозирование динамики поведения и симуляцию сценариев. Непрерывный анализ интенсивности позволит выявить эффекты лояльности и насыщения, способствуя повышению отдачи от маркетинговых вложений в цифровой экономике. Диагностика временной структуры отклика определит возможности адаптивных стратегий, четко очерчивая границы воздействия модели на оптимизацию таргетинга.

**Ключевые слова:** нейронные процессы Хоукса; механизмы внимания; моделирование пользовательского поведения; цифровые платформы; оценка интенсивности; маркетинговые стратегии; временная динамика; векторные представления; оптимизация таргетинга.

**Введение.** Несмотря на то, что ежегодные маркетинговые расходы в музыкальной индустрии превышают 200 миллиардов долларов США [1], оценка эффективности действий в этой области остается сложной задачей. Современные цифровые платформы сталкиваются с возрастающими ограничениями на сбор и обработку данных, обусловленными регуляторными мерами, такими как закон «О персональных данных». В результате маркетинговые модели, полагающиеся на полный доступ к данным, становятся менее применимыми, что требует разработки подходов, способных эффективно работать с ограниченными или агрегированными данными. В контексте музыкальных потоковых платформ данные о прослушиваниях часто включают только временные метки событий (например, факт прослушивания) и ограниченный набор категориальных переменных, таких как источник прослушивания (плейлист, радио, алгоритмическая рекомендация), возрастная группа, страна или время суток. Важность учёта не только факта последнего прослушивания, но и истории взаимодействий слушателей с различными характеристиками подчёркивается в исследованиях, посвящённых анализу пользовательского поведения на цифровых платформах [2].

Согласно эмпирическим исследованиям, пользователи классифицируются по уровню разнообразия потребления в две основные группы: «специалисты» и «генералисты». Специалисты демонстрируют низкое разнообразие, фокусируясь на узкой группе похожих произведений с высокой схожестью их векторных представлений. Генералисты, напротив, демонстрируют высокое разнообразие, охватывая широкий спектр сегментов каталога, с низкой схожестью между потребляемыми треками [3-5].

Кроме того, анализ динамики платформ выявляет третью группу – интенсивных пользователей (*heavy users*), у которых эффекты платформенного потребления усиливаются со временем, повышая сходство поведения, объем взаимодействий и повторяемость произведений через плейлисты и рекомендации, что усиливает самовозбуждение. Существующие различия подчеркивают необходимость адаптивных стратегий анализа, учитывающих краткосрочную релевантность и долгосрочное разнообразие [4].

Целью исследования является разработка и анализ модели для оценки динамики спроса на музыкальных потоковых платформах с учетом ограниченных данных и частичного наблюдения, с акцентом на влияние истории взаимодействий, алгоритмических рекомендаций и маркетинговых действий. Предлагаемая модель учитывает эффекты самовозбуждения и ингибирования, с целью повышения точности прогнозирования динамики пользовательского поведения и оптимизации маркетинговых стратегий.

### **Разработка эконометрической модели**

Разрабатываемая модель может рассматриваться как модель повторяющегося дискретного выбора, в которой потребители оценивают товары на основе внутренних и внешних источников информации, динамически обновляющихся предпочтений [6-7]. Согласно такой модели выбор потребителем в каждый момент времени  $t$  дискретен, а спрос  $D_t$  является агрегированным счётчиком событий [8]. Факт потребления происходит, если функция полезности положительна для потребителя, и каждый период часть потребителей  $\alpha < 1$  решает, если стоит продолжать потребление данного продукта или нет. Тогда существует следующая динамика [9]:

$$D_{t+1} = (1 - \alpha)D_t + \alpha \cdot \Delta D_t^*, \quad (1)$$

где:

$D_t$  – спрос, агрегированный счётчик событий в момент  $t$ ;

$\alpha$  – доля потребителей, пересматривающих свой выбор в каждом периоде от общего количества потребителей, где  $0 < \alpha < 1$ ;

$\Delta D_t^*$  – изменение спроса, зависящее от эндогенных факторов.

Предположим, что выбор описывается функцией плотности вероятности  $f(D_{<t}, R_{<t}, A_t)$ , определённой на множестве внимания (аналогичном бюджетному множеству, ограниченному временем/вниманием). Ожидаемый спрос определяется как:

$$E[D_t] = \int D_t f(D_{<t}, R_{<t}, A_t) dD_t, \quad (2)$$

где:

$E[D_t]$  – ожидаемый спрос в момент времени  $t$ ;

$D_{<t}$  – история потребления до момента  $t$ , эндогенно сформированные предпочтения, привычки, путь пользователя;

$R_{<t}$  – алгоритмические рекомендации, зависящие от агрегированной истории  $D_{<t}$ ;

$A_t$  – маркетинговые действия (включения в плейлисты, чарты, альбомы).

В рамках анализа событиями рассматриваются акты потребления цифрового контента, отдельные факты воспроизведения трека пользователем. Каждое событие характеризуется как история реализации данного процесса  $H_t$ , как множество событий в интервале  $[0, t)$ , а именно  $H_t = \{(t_i, k_i) | t_i < t\}$ , где  $t_i$  соответствует времени наступления события  $i$ , а  $k_i$  – типу события (например, радио, чарты, персональная коллекция, плейлист, рекомендации).

Для каждого типа источника  $k$  вводится функция интенсивности  $\lambda_k$ , описывающая мгновенную вероятность появления нового события данного типа в момент времени  $t$  при условии всей истории системы до этого момента, таким образом функция спроса может быть представлена как:

$$D(t) = \sum_k N_k(t), \quad (3)$$

где:

$D_t$  – спрос, агрегированный счётчик событий в момент  $t$ ;

$N_k$  – счётный процесс для класса  $k$ .

### Существующие модели и их ограничения

Традиционные модели маркетингового анализа, такие как счётные модели, модели выживания и их вариации, включая BG/NBD (Beta-Geometric/Negative Binomial Distribution) и Pareto/NBD (Pareto/Negative Binomial Distribution), часто используются для прогнозирования повторных покупок и анализа временных рядов событий [10]. Дискретно-временные подходы, предполагают независимость событий или постоянную интенсивность, что ограничивает их способность учитывать кластеризацию событий во времени и динамические взаимовозбуждающие эффекты. Такие упрощенные предположения о сложной динамике точечных процессов значительно ограничивают практическую применимость, игнорируя влияние предшествующих взаимодействий.

Модели процессов Хоукса позволяют учитывать такие эффекты, но в классической интерпретации предполагают, что все прошлые события должны оказывать положительное влияние на возникновение текущих событий [11-12].

Хотя большинство приложений процессов Хоукса сосредоточено исключительно на моделировании возбуждения (то есть возникновение одного события увеличивает

вероятность будущих событий), они также могут использоваться для описания ингибирования (то есть возникновение одного события снижает вероятность будущих событий), что не менее актуально для анализа взаимного влияния типов прослушивания музыкальных произведений для разной аудитории на цифровых потоковых платформах. Динамика спроса на музыкальные произведения может формироваться последовательностью событий, которые могут как стимулировать, так и ингибировать последующие прослушивания через механизмы самовозбуждения [13].

Моделирование большого количества категориальных переменных в рамках традиционных процессов Хоукса без применения архитектур нейронных сетей оказывается малоэффективным, поскольку требует явного учета всех возможных взаимодействий, что приводит к экспоненциальному росту параметров модели и потере обобщающей способности [14]. Такие подходы, основанные на фиктивных индикаторах или простых параметрических функциях, не захватывают долгосрочные последовательности событий, которые для каждого пользователя могут быть индивидуальными и зависеть от комбинаций факторов, таких как демографические характеристики и контекст прослушивания. В результате игнорируются сложные зависимости, где события, разделенные значительными временными интервалами, могут оказывать существенное влияние на текущую интенсивность, что приводит к недооценке эффектов самовозбуждения [15].

Для преодоления указанных ограничений в данной статье предлагается моделирование взаимного влияния различных типов прослушиваний с использованием архитектур нейронных сетей, в частности механизмов внимания, подобных тем, что применяются в моделях трансформеров, интегрированных в структуру процессов Хоукса. Этот подход обеспечивает учёт сложных взаимосвязей между типами прослушиваний и демографическими характеристиками пользователей, обеспечивая адаптивное выделение значимых событий из истории, что улучшает точность прогнозирования динамики.

В рамках анализа пользовательских данных на цифровых платформах ключевое затруднение состоит в том, что доступные наблюдения отражают лишь часть процесса взаимодействия: фиксируются только отдельные события (например, акты прослушивания), тогда как скрытые стимулы и алгоритмические воздействия остаются ненаблюдаемыми. Это приводит к проблеме частичного наблюдения и динамической эндогенности: вероятность нового события зависит от истории прошлых взаимодействий, а сами маркетинговые и алгоритмические воздействия формируются на их основе. В таких условиях стандартные дискретно-временные модели не позволяют корректно оценить временную структуру отклика. Использование процессов Хоукса позволяет формально учесть эти взаимные зависимости и интерпретировать параметры интенсивности как экономически осмысленные эффекты самовозбуждения (лояльности) и ингибирования (насыщения) спроса.

Аналогичные методы, основанные на использовании векторных представлений в рамках процессов Хоукса, находят применение в смежных областях. Например, моделирование событий в медицинских диагностических траекториях или социальных сетях, где события представляются в векторной форме, что обеспечивает баланс между интерпретируемостью и гибкостью модели [16-18]. Метод проверки НоТРР для моделирования длинных горизонтов прогнозирования также подчёркивает важность использования векторных представлений [19].

**Материалы и методы.** Интенсивность  $\lambda_k$  моделируется с помощью процесса Хоукса с расширением на маркированные события и нелинейность:

$$\lambda_k(t | H_t) = \sigma(\mu_k + \sum_{t_i < t} \phi_{k,k_i}(t - t_i)), \quad (6)$$

где:

$\lambda_k(t|H_t)$  – условная интенсивность для типа события  $k$  в момент  $t$ , учитывающая историю  $H_t$ ;

$\sigma(\cdot)$  – нелинейная функция активации, обеспечивающая положительность;

$\mu_k$  – базовая интенсивность для типа события  $k$ , представляющая фоновую скорость событий, независимую от прошлых событий, характеризует экзогенный уровень спроса, т. е. вероятность потребления без учета прошлых событий. Экономически это соответствует органическому спросу на контент данного типа (например, постоянным слушателям радио);

$\phi_{k_i}$  – функция воздействия (ядро самовозбуждения), отражает внутреннюю инерцию потребления, как событие типа  $k_i$  в момент  $t_i$  влияет на интенсивность события типа  $k = k_i$  в будущем, например, если трек был воспроизведён через тот же источник, вероятность повторного воспроизведения возрастает. Экономически это соответствует эффектам лояльности или привычки;

$\phi_{k,k_i}(\cdot)$  – функция воздействия (ядро перекрестного возбуждения), описывающая, как событие типа  $k_i$  в момент  $t_i$  влияет на интенсивность события типа  $k \neq k_i$  в будущем, описывают межканальные перетоки спроса: например, событие «прослушивание через радио» повышает вероятность события «добавление трека в коллекцию». Эти эффекты интерпретируются как переходы от пассивного к активному потреблению, или как воздействие маркетингового и алгоритмического продвижения на будущий спрос;

$t_i$  – время наступления прошлого события  $i$ ;

$k$  – тип текущего события;

$k_i$  – тип прошлого события  $i$ .

**Методология исследования.** Методология интегрирует идеи из недавних работ по нейронным процессам Хоукса в анализе пользовательского поведения в стриминге и цифровом контенте, обеспечивая баланс интерпретируемости и гибкости в условиях ограниченных данных [16-17].

Предлагаемая модель представляет собой смешанный подход, где события описываются в виде векторов векторных представлений и производится суммирование этих векторов для вычисления интенсивности для сохранения аддитивности влияний, аналогично работе [17]. В то же время структура моделирования интенсивности основана на интерпретируемых трансформерных процессах Хоукса, что позволяет учитывать сложные взаимодействия между типами событий без потери интерпретируемости, аналогично работам [16; 20].

Методология исследования включает последовательный алгоритм, состоящий из следующих этапов, направленных на обработку данных, построение модели и анализ результатов:

1. Описание данных и их предобработки, включая нормализацию временных наблюдений и преобразование категориальных переменных в векторные представления.
2. Моделирование интенсивности с трансформерной модели точечного процесса Хоукса с расширением на маркированные события и интеграцией механизмов внимания для моделирования функции воздействия.
3. Разделение данных на выборки для обучения и валидации, с последующей оценкой качества модели через остаточный анализ и стандартные метрики регрессии.

4. Реализация раздельного обучения для категорий событий и механизма внимания, включая анализ ядер воздействия и количественную оценку их суммарного влияния.

5. Оценка важности независимых переменных через пост-обработку весов внимания для интерпретации вкладов характеристик в динамику интенсивности.

**Описание данных.** Настоящее исследование охватывает около 300 музыкальных произведений, представленных в виде временных рядов, прослушивания композиций разделены на данные 40000 пользователей, с временными метками за период с сентября по октябрь 2024 года.

Фактические временные наблюдения нормализованы путём вычисления времени с момента выпуска трека, разделённого на стандартное отклонение, чтобы уменьшить асимметрию данных. Категориальные переменные, такие как источник прослушиваний; характеристики, такие как возрастная группа, пол и время дня (вычисляемая как утро, день, вечер или ночь на основе временной метки), преобразуются в векторные представления.

### **Точечные процессы Хоукса на базе трансформерных моделей**

Традиционные параметрические формы  $\phi_{k,k_i}$  (например, экспоненциальное затухание) ограничивают гибкость, что приводит к недооценке нелинейных зависимостей в реальных экономических данных [21].

Для преодоления ограничений традиционных моделей аналогично [16-17; 22] предлагается интерпретируемый трансформерный процесс Хоукса, который использует механизмы внимания для захвата сложных взаимодействий в социальных сетях. В статье эти идеи адаптируются для маркетингового контекста, где события – это факты прослушивания композиции, а встраивания позволяют моделировать характеристики события в низкоразмерном пространстве.

Функция воздействия  $\phi_{k,k_i}(t-t_i)$  моделируется как трансформерная нейронная сеть:

$$\phi_{k,k_i}(\tau) = f(t-t_i, e_k, e_{k_i}, a'_i; \theta), \quad (7)$$

где:

$f$  – трансформерный энкодер с механизмами само-внимания, параметризованный  $\theta$ ;

$\tau = t - t_i$  – временной лаг между событиями;

$e_k, e_{k_i}$  – векторные встраивания типов событий  $k$  и  $k_i$ , обучаемые параметры;

$a_i$  – независимая переменная события  $k_i$ , характеристика прослушивания;

$a'_i$  – векторные встраивания независимых переменных, обучаемые параметры.

$\theta$  – параметры трансформерного энкодера, оцениваемые на данных.

В реализации модели энкодер принимает исходный тип событий в виде низкоразмерных векторов и временных меток. Для каждого типа  $i$  модель вычисляет выходной вектор энкодера с использованием временного встраивания (positional encoding), встраиваний типов и независимых переменных, затем применяет декодер интенсивности для получения  $\lambda_k$ . Параметры оцениваются путём максимизации логарифма правдоподобия  $L$  на исторических данных о пользовательских взаимодействиях, что позволяет симулировать сценарии экзогенных воздействий для оптимизации маркетинга [23].

Максимизация  $L$  относительно параметров проводится с использованием градиентных методов, таких как AdamW, с регуляризацией L2 для предотвращения переобучения.

### Разделение данных на выборки

С практической точки зрения, данные состоят из  $n$  последовательностей событий, каждая из которых рассматривается как реализация стохастического процесса Хоукса, где история событий обозначается как  $H_t = \{(t_i, k_i, a_i) | t_i < t\}$ , с  $t_i$  – временем события,  $k_i$  – его типом и  $a_i$  – его характеристиками. Наблюдения разбиваются на две выборки: выборку для обучения и для валидации, на основе которой оцениваются параметры модели. Это разделение обеспечивает оценку обобщающей способности модели на не доступные при обучении данные.

### Оценка качества модели

Точность оценки параметров интенсивности имеет ключевое значение, поскольку она напрямую определяет способность модели захватывать временную динамику взаимодействий пользователей и влияние экзогенных факторов.

Для формальной проверки качества модели точечного процесса мы используем остаточный метод на основе теоремы Папангелу о случайном изменении времени. Кумулятивная функция интенсивности выражается как:

$$\Lambda(t) = \int_0^t \lambda_k(s | H_s) ds, \quad (8)$$

где:

$\Lambda(t)$  – кумулятивная интенсивность до момента  $t$ , представляющая ожидаемое число событий в интервале  $[0, t]$ ;

$\lambda_k(s | H_s)$  – условная интенсивность для типа события  $k$  в момент  $s$ , учитывающая историю  $H_s$ ;

$H_s$  – история событий до  $s$ , включая времена  $t_j$ , типы  $k_j$  и независимые переменные  $a_j$ ;

$ds$  – дифференциал времени.

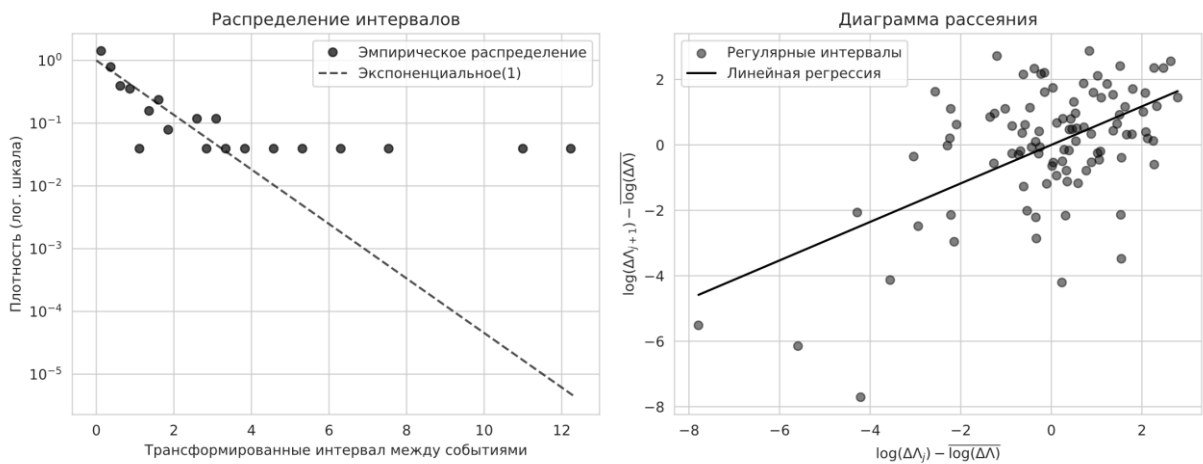
Соответствующий трансформированный счётный процесс имеет вид  $N(\Lambda^{-1}(u))$ . Если модель точно захватывает характеристики данных о взаимодействиях пользователей, то оценённые времена прибытия событий  $\hat{t}_i = 1, \dots, N(T)$  должны приводить к интервальным остаткам:

$$\{\Lambda(\hat{t}_1), \Lambda(\hat{t}_2) - \Lambda(\hat{t}_1), \Lambda(\hat{t}_3) - \Lambda(\hat{t}_2), \dots\} = \{\tau_1^*, \tau_2^* - \tau_1^*, \tau_3^* - \tau_2^*, \dots\}, \quad (9)$$

где эти остатки ожидаемо следуют экспоненциальному распределению с единичной интенсивностью:  $\tau_i^* \sim \exp(1)$ .

Рисунок 1 демонстрирует результаты теста на соответствие эмпирического распределения интервалов между событиями теоретическому экспоненциальному распределению с параметром масштаба равным единице, а также анализ последовательной зависимости в логарифмической шкале. Эмпирическая плотность (точки) в целом согласуется с теоретической плотностью экспоненциального распределения (пунктирная линия), за исключением умеренного отклонения в правом хвосте. Вероятно, это обусловлено наличием высокоактивных слушателей, создающих непропорционально высокий отклик. Диаграмма рассеяния нормализованных логарифмированных последовательных интервалов выявляет слабую положительную зависимость между последовательными событиями (коэффициент корреляции Спирмена

равен 0,121 со статистической значимостью  $p < 0.05$ ), что указывает на отсутствие сильной автокорреляции.



**Рис. 1. Усредненные диаграммы распределения интервалов и рассеяния логарифмированных последовательных интервалов между событиями**

Для количественной оценки предсказательной точности модели мы вычисляем стандартные метрики регрессии для обучающей и валидационной выборки предсказаний интенсивности  $\lambda_k(t | H_t)$ . Результаты приведены в таблице 1.

**Таблица 1. Метрики качества предсказания интенсивности модели**

Метрика	Обучающая выборка	Валидационная выборка
MSE, среднеквадратичная ошибка	0.023	0.045
MAE, средняя абсолютная ошибка	0.112	0.156
$R^2$ , коэффициент детерминации	0.912	0.874
RMSE, корень из MSE	0.152	0.212
MAPE, средняя абсолютная процентная ошибка	8.45%	12.3%

Оценка предсказательной точности модели, демонстрирует высокую эффективность: на обучающей выборке коэффициент детерминации  $R^2 = 0.912$  указывает на объяснение 91.2% дисперсии интенсивности, с низкими ошибками  $MSE = 0.023$  и  $MAPE = 8.45\%$ ; на валидационной выборке значения слегка ухудшаются ( $R^2 = 0.874$ ,  $MSE = 0.045$ ,  $MAPE = 12.3\%$ ), что свидетельствует об адекватной обобщающей способности и минимальном переобучении, подтверждая применимость модели для симуляции маркетинговых воздействий в условиях ограниченных данных.

### **Раздельное обучение категорий и механизм внимания**

Для повышения интерпретируемости и точности моделирования взаимодействий между категориями событий модель реализует раздельное обучение интенсивностей для каждого типа событий. Этот подход вдохновлён работами [16-17], где подчёркивается важность разделения параметров модели для различных категорий событий, чтобы учесть их уникальные характеристики, такие как различные временные зависимости или влияние экзогенных факторов.

Раздельное обучение реализовано через цикл по типам событий, где для каждого типа  $k$  все входные типы событий временно заменяются на  $k$ , чтобы вычислить

интенсивность  $\lambda_k$ . Это позволяет модели фокусироваться на влиянии истории  $H_t$  на конкретный тип  $k$ , избегая смешивания параметров между категориями, как в традиционных моделях Хоукса.

Для каждого исторического события  $(t_i, k_i, a_i)$ , входной вектор формируется как конкатенация временного встраивания  $t_i$ , типового встраивания  $e_{k_i}$  и атрибутивного встраивания  $a'_i$ .

В оригинальной работе [24] вводится концепция, где запрос  $Q$ , ключ  $K$  и значение  $V$  представляют собой матрицы, полученные путём линейных проекций входных встраиваний. Запрос  $Q$  определяет, на что фокусируется внимание, ключ  $K$  используется для вычисления сходства с запросом, а значение  $V$  агрегируется с весами внимания для формирования выхода. В данной модели  $Q$  и  $K$  формируются непосредственно из входных встраиваний для сохранения инвариантности к временным сдвигам и вычисляются как:

$$Q = [t, e_k, 0_{d_m}], \quad K_i = V_i = [t_i, e_{k_i}, a'_i], \quad (10)$$

где:

$0_{d_m}$  – нулевой вектор независимых переменных, обеспечивающий фокус на типе  $k$  размерности  $d_m$ ;

$e_k, e_{k_i}$  – встраивания типов событий  $k$  и  $k_i$ , обучаемые параметры;

$t, t_i$  – время текущего и прошлого события  $i$ ;

$a'_i$  – встраивания независимых переменных, обучаемые параметры.

Веса внимания для текущего события по отношению к прошлому событию  $i$  (где  $t < t_i$ ) вычисляются через масштабированное внимание точечного продукта:

$$S = \text{soft max} \left( \frac{QK_j}{\sqrt{d_k}} \right), \quad (11)$$

где:

$S$  – вес внимания, отражающий влияние текущего события на событие  $i$ ;

$Q$  – матрица запросов, полученная преобразованием встраиваний и их весов  $W^Q$  где  $W^Q \in R^{L \times d_K}$ ;

$K$  – матрица ключей, полученная преобразованием встраиваний и их весов  $W^K$  где  $W^K \in R^{L \times d_K}$ ;

$L$  – длина последовательности, используемая для обучения модели;

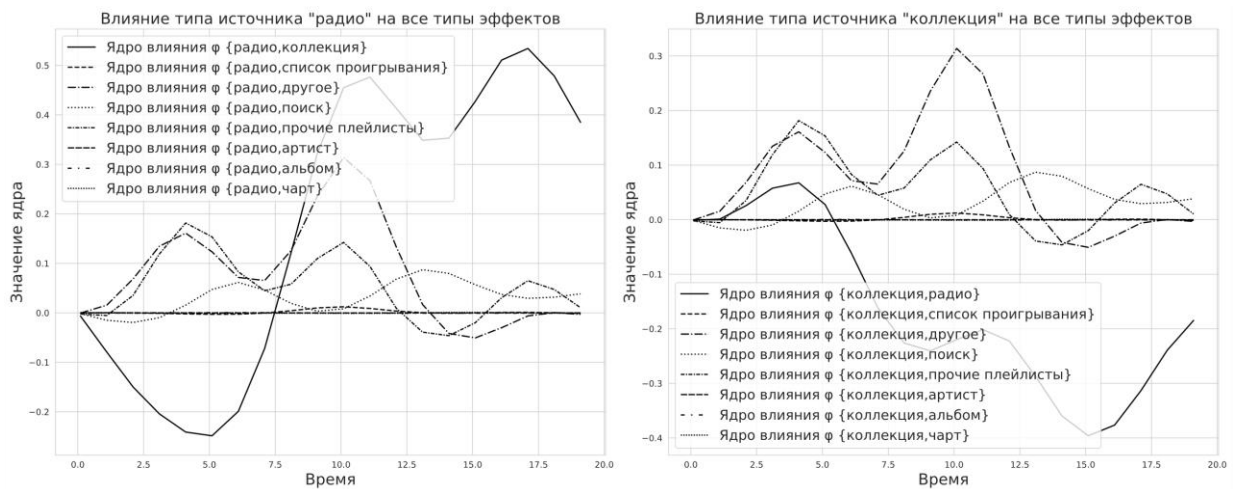
$d_k$  – размерность ключей, обеспечивающая масштабирование для стабильности;

$\text{softmax}$  – функция нормализации, гарантирующая, что сумма всех значений не превысит единицу.

Рисунок 2 иллюстрирует функции воздействия (ядра возбуждения)  $\phi_{k,k_i}(t)$  для одного случайного произведения для различных пар типов событий, где  $k$  – тип-причина (источник), а  $k_i$  – тип-эффект (целевое взаимодействие):

- «радио» – моделирует экзогенный фактор, такой как алгоритмическая рекомендация;
- «коллекция» – моделирует эндогенный переход к лояльности.

Анализ ядер выявляет дифференцированные динамические паттерны, отражающие поведенческие механизмы пользователей. Так, например, взаимодействие «радио» – «коллекция» вначале демонстрирует ингибирующее влияние (отрицательные значения в интервале  $t \approx 0-8$ ), однако со временем наблюдается переход к положительному эффекту с пиком около  $t \approx 15$  единиц времени, указывающий на задержанное самоусиление: пользователи не сразу добавляют понравившиеся произведения в свои коллекции. Остальные ядра показывают или минимально-положительное воздействие («поиск»), указывающий на исследовательское поведение после того, как композиция услышана на радио, или близкое к нулю. Исключение составляет тип «прочие плейлисты» с отрицательным (ингибирующим) эффектом, что согласуется с теорией замещения в цифровом маркетинге [25].



**Рис. 2. Ядра воздействия  $\phi_{k,k_i}(t)$  для различных пар типов событий на примерах событий причин «радио» и «коллекция»**

Влияние типа «коллекция» показывает, что добавление в коллекцию снижает вероятность повторных взаимодействий через другие каналы (поиск, плейлисты, чарты), что кажется обоснованным: пользователь переходит к прямому доступу через личную библиотеку, за исключением типа «поиск», отражающий целенаправленное поведение, пользователь ищет уже знакомый контент для добавления или повторного прослушивания.

**Результаты.** Суммарное влияние типа-причины  $k_i$  на все типы-эффекты  $k \neq k_i$  определяется как интеграл ядер возбуждения по времени для всех пар, исключая авто-воздействие:

$$I_k = \sum_{k \neq k_i} \int_0^{\infty} \phi_{k,k_i}(\tau) d\tau, \quad (12)$$

где:

$I_k$  – суммарное воздействие типа  $k$  на все другие типы  $k_i$ ;

$\phi_{k,k_i}(\tau)$  – ядро возбуждения от события типа  $k$  на тип  $k_i$  с лагом  $\tau$ .

Оптимальный момент вмешательства (наиболее эффективный лаг  $\tau^*$ ) для типа-причины  $k$  определяется как:

$$\tau_k^* = \arg \max_{\tau} \sum_{k \neq k_i} \phi_{k,k_i}(\tau), \quad (13)$$

где:

$\tau_k^*$  – оптимальный временной лаг для максимальной эффективности воздействия типа  $k$ , соответствует пику суммарного мгновенного воздействия на все типы-эффекты;  
 $\arg \max_{\tau}$  – аргумент, максимизирующий суммарное ядро.

**Таблица 2. Количественная оценка суммарного влияния типов-причин на все типы-эффекты**

Тип причины, $k_i$	Общее влияние $I_{k_i}$	Оптимальное время $\tau_{k_i}^*$	Пик влияния $\sum_{k \neq k_i} \phi_{k,k_i}(\tau_{k_i}^*)$
радио	2.809	14.1	0.740
прочие плейлисты	2.670	13.1	1.009
альбом	1.377	12.4	0.986
чарт	1.275	11.9	0.972
артист	0.949	13.3	0.855
список проигрывания	0.462	12.9	0.556
другое	0.159	12.2	0.374
поиск	- 0.311	11.7	0.289
коллекция	- 1.394	12.5	0.134

Наибольшее суммарное влияние  $I_{k_i} = 2.809$  наблюдается для типа «радио», что подчёркивает роль экзогенных факторов в генерации каскадных эффектов на платформе. Оптимальное время  $\tau_{k_i}^* \approx 12-14$  указывает на задержанный пик воздействия, что соответствует эмпирическим паттернам в маркетинге, где пользователи реагируют не мгновенно, а после периода принятия решения. Второе по величине влияние  $I_{k_i} = 2.670$  демонстрирует тип «прочие плейлисты», предположительно вследствие эффекта замещения: альтернативные рекомендации конкурируют с основными каналами, снижая их эффективность. Типы «альбом»  $I_{k_i} = 1.377$  и «чарт»  $I_{k_i} = 1.275$  показывают средние положительные эффекты с близкими оптимальными временами (12.4 и 11.9 соответственно), что свидетельствует о их роли в стимулировании группового потребления. Отрицательные значения суммарного влияния наблюдаются для «поиск»  $I_{k_i} = -0.311$  и «коллекция»  $I_{k_i} = -1.394$ , что интерпретируется как ингибирование: эти типы подавляют другие взаимодействия, переключая пользователей на прямой доступ (например, после добавления в коллекцию снижается нужда в поиске или радио). В целом, асимметрия влияний подтверждает практическую ценность модели: симуляция на основе этих метрик позволяет оптимизировать временные параметры рекламных кампаний.

Для оценки вклада независимых переменных (таких как гендерные или возрастные характеристики пользователей) вводится их пост-обработка на основе весов внимания. Этот подход позволяет визуализировать и количественно оценивать влияние атрибутов событий на функцию воздействия  $\phi_{k,k_i}(\tau)$ . В отличие от модели TransFeat-TPP, где интерпретируемость достигается через отдельный модуль Fi-SAN для вычисления важности независимых переменных без учёта типов событий, реализация учитывает взаимодействие между типами событий и атрибутами: атрибуты встраиваются в энкодер (через  $a'_i$ ), как следствие общий вклад характеристик может снижаться или подавляться доминирующим влиянием типов. В модели осуществляется конкатенация векторов выхода энкодера и внутриатрибутивных представлений, что позволяет полностью

учитывать атрибуты в процессе обучения, интегрируя их в единую последовательность представлений для последующего внимания и предсказания [20].

Математически, важность компоненты  $c$  независимых переменных (элемента вектора  $a_i$ ) оценивается как средние веса внимания в модуле внутриатрибутивного внимания, реализованном как сеть само-внимания:

$$FI_c(a_i) = \frac{1}{H} \sum_{h=1}^H \text{soft max}(W_{h,c} a'_i + b_{h,c}), \quad (14)$$

где:

$FI_c(a_i)$  – важность компоненты  $c$  для атрибутов события  $i$ ;

$a_i$  – вектор независимых переменных для события  $i$ ;

$a'_i$  – встраивания независимых переменных, обучаемые параметры;

$H$  – количество голов внимания в сети само-внимания (в данной реализации  $H = 8$  для внутриатрибутивного модуля);

$W_{h,c}$  – веса линейного слоя для головы  $h$  и компоненты  $c$ , обучаемые параметры;

$b_{h,c}$  – свободный член (смещение) в линейном преобразовании для головы внимания  $h$  и компоненты  $c$ , обучаемые параметры.

Глобальная важность компоненты  $c$  по всей последовательности событий  $H_t = \{(t_i, k_i, a_i) | t_i < t\}$  вычисляется как среднее по всем событиям:

$$FI_c(H_t) = \frac{1}{N} \sum_{i=1}^N FI_c(a_i), \quad (15)$$

где:

$FI_c(H_t)$  – глобальная важность компоненты  $c$  (усреднённая по весам голов внимания и событиям);

$N$  – количество событий в истории  $H_t$ .

Данная формула расширяет стандартные механизмы внимания [24], обеспечивая интерпретируемость, аналогично модулю Fi-SAN. В предложенной модели атрибуты событий конкатенируются с временными и типовыми встраиваниями, что, с одной стороны, увеличивает вычислительную сложность задачи, с другой стороны, позволяет обучаемым весам эффективно разделять вклад каждой компоненты. Кроме того, в модели осуществляется конкатенация векторов выхода энкодера и внутриатрибутивных представлений, что позволяет полностью учитывать атрибуты в процессе обучения, интегрируя их в единую последовательность представлений для последующего внимания и предсказания.

Рисунок 3 отображает пример оценки важности компонент независимых переменных для случайного произведения, демонстрируя их вклад в функцию воздействия  $\phi_{k,k_i}(\tau)$ , что позволяет симулировать влияние экзогенных факторов на интенсивность событий в маркетинговых кампаниях.

Независимые переменные могут использоваться совместно с типами событий для более точной оценки параметров интенсивности, что позволяет симулировать сложные маркетинговые сценарии.

Анализ показывает, что для группы «Пол» наибольший вклад в интенсивность событий вносит категория «женский» (0.141), что может указывать на более высокую

отзывчивость женской аудитории к экзогенным факторам. В группе «Часть дня» доминирует «Вечер» (0.172), подчёркивая пик пользовательской активности в вечерние часы, что практически для планирования временных рамок пиков рекламных кампаний. Наконец, в возрастной группе наибольшее внимание приходится на «35-44» (0.181), с постепенным снижением для старших категорий (до 0.094 для «55+»), что помогает уточнить фокус аудиторию для маркетинга.

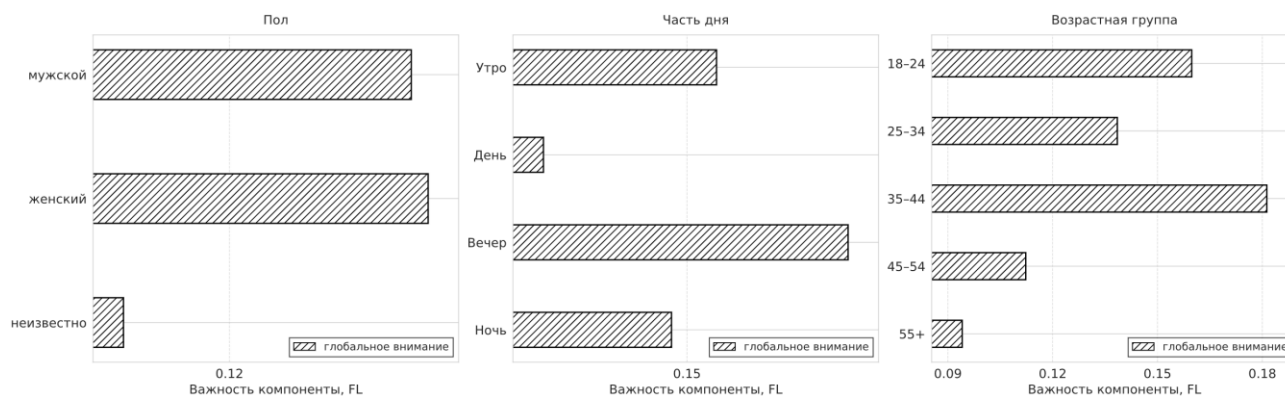


Рис. 3. Результаты оценки важности характеристик независимых переменных

**Обсуждение результатов.** Полученные результаты демонстрируют эффективность предложенной модели на основе процессов Хоукса с механизмами внимания в оценке параметров интенсивности, что позволяет учитывать сложные взаимодействия между типами событий и независимыми переменными в контексте цифровых платформ. В частности, анализ ядер воздействия подчеркивает асимметрию влияний: экзогенные факторы, такие как рекомендации или внешние стимулы, проявляют положительное возбуждение с задержкой, в то время как эндогенные взаимодействия, связанные с лояльностью пользователей, часто приводят к ингибированию последующих событий. Особое внимание следует уделить совместному использованию независимых переменных с типами событий. Независимые переменные, такие как демографические характеристики пользователей (возрастная группа, пол) или контекстные факторы (время суток, географическое положение), интегрируются в модель через векторные представления, что позволяет учитывать их взаимодействие с типами событий в пространстве векторных представлений. Такой подход обеспечивает адаптивное выделение значимых комбинаций: например, влияние типа события «рекомендация» усиливается для определенных возрастных групп, что проявляется в повышенной интенсивности последующих взаимодействий. Кроме того, модель вычисляет временную составляющую, что дает возможность симулировать сценарии, в которых временные параметры кампаний (например, момент запуска рекламы) оптимизируются для максимального воздействия.

Качество модели подтверждается метриками, такими как низкие значения MSE и MAE на валидационной выборке, а также соответствием остатков экспоненциальному распределению, что указывает на адекватный учет кластеризации событий. Однако наблюдаемое отклонение в хвостах распределения может быть связано с гетерогенностью пользователей, требующей дальнейшей стратификации данных.

**Заключение.** Разработанная модель на основе процессов Хоукса с трансформерными механизмами внимания представляет собой эффективный инструмент для оценки параметров интенсивности в условиях ограниченного доступа к данным на цифровых платформах. Полученные результаты демонстрируют способность модели захватывать эффекты самовозбуждения и ингибирования, а также учитывать взаимодействия типов событий и независимых переменных с временной динамикой, что

позволяет симулировать воздействие экзогенных факторов, таких как маркетинговые кампании или активность целевых групп пользователей. Анализ показал, что эффекты самовозбуждения (ассоциируемые с лояльностью) и ингибирования (связанные с насыщением) спроса имеют экономический смысл и могут быть количественно оценены через параметры интенсивности.

Практическая ценность заключается в оптимизации стратегий таргетинга: на основе оцененных ядер воздействия и важности переменных возможна разработка адаптивных кампаний, учитывающих оптимальные временные лаги для максимальной эффективности. Это способствует повышению отдачи от маркетинговых вложений в цифровой экономике, где традиционные подходы ограничены регуляторными барьерами.

Дальнейшие исследования могут быть направлены на разработку методов долгосрочного прогнозирования на основе предложенной модели, а также симуляции контрафактических сценариев.

### Список литературы

1. Xu L., Duan J. A., Whinston A. Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion // *Management Science. Institute for Operations Research and the Management Sciences (INFORMS)*, 2014. Т. 60, № 6. С. 1392–1412.
2. Anderson A. и др. Algorithmic Effects on the Diversity of Consumption on Spotify // *Proceedings of The Web Conference 2020. Taipei Taiwan: ACM*, 2020. С. 55–65.
3. Datta H., Knox G., Bronnenberg B. J. Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery // *Marketing Science*. 2018. Т. 37, № 1. С. 5–21.
4. Morris J. W., Powers D. Control, curation and musical experience in streaming music services: 2 // *Creative Industries Journal*. 2015. Т. 8, № 2. С. 106–122.
5. Anas A. Discrete choice theory, information theory and the multinomial logit and gravity models // *Transportation Research Part B: Methodological*. 1983. Т. 17, № 1. С. 13–23.
6. Ben-Akiva M. и др. Modeling Methods for Discrete Choice Analysis // *Marketing Letters*. 1997. Т. 8, № 3. С. 273–286.
7. Leake C. Discrete Choice Theory of Product Differentiation // *Journal of the Operational Research Society*. 1995. Т. 46, № 4. С. 543–543.
8. Cowan R., Cowan W., Peter Swann G. M. Waves in consumption with interdependence among consumers // *Canadian J of Economics*. 2004. Т. 37, № 1. С. 149–177.
9. Fader P. S., Hardie B. G. S., Lee K. L. "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model // *Marketing Science. Institute for Operations Research and the Management Sciences (INFORMS)*, 2005. Т. 24, № 2. С. 275–284.
10. Hawkes A. G. Spectra of some self-exciting and mutually exciting point processes // *Biometrika. Oxford University Press (OUP)*, 1971. Т. 58, № 1. С. 83–90.
11. Isham V., Westcott M. A self-correcting point process // *Stochastic Processes and their Applications. Elsevier BV*, 1979. Т. 8, № 3. С. 335–347.
12. Eichler M., Dahlhaus R., Dueck J. Graphical Modeling for Multivariate Hawkes Processes with Nonparametric Link Functions. arXiv, 2016.
13. Du N. и др. Recurrent Marked Temporal Point Processes: Embedding Event History to Vector // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM*, 2016. С. 1555–1564.
14. Mei H., Eisner J. The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process. arXiv, 2016.
15. Meng Z. и др. Interpretable Transformer Hawkes Processes: Unveiling Complex Interactions in Social Networks. arXiv, 2024.

16. Zhao Y., Engelhard M. Balancing Interpretability and Flexibility in Modeling Diagnostic Trajectories with an Embedded Neural Hawkes Process Model. arXiv, 2025.
  17. Zuo S. и др. Transformer Hawkes Process. arXiv, 2020.
  18. Karpukhin I., Dereka S., Kolesnikov S. EXACT: How to train your accuracy // Pattern Recognition Letters. 2024. Т. 185. С. 23–30.
  19. Meng Z. и др. TransFeat-TPP: An Interpretable Deep Covariate Temporal Point Processes // Frontiers in Artificial Intelligence and Applications / под ред. Endriss U. и др. IOS Press, 2024.
  20. Kirchner M. An estimation procedure for the Hawkes process. arXiv, 2015.
  21. Zuo S. и др. Transformer Hawkes Process. arXiv, 2020.
  22. Ogata Y. Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes // Journal of the American Statistical Association. JSTOR, 1988. Т. 83, № 401. С. 9.
  23. Vaswani A. и др. Attention Is All You Need: arXiv:1706.03762. arXiv, 2023.
  24. Deutsch I., Ross G. J. Estimating Product Cannibalisation in Wholesale using Multivariate Hawkes Processes with Inhibition. arXiv, 2022.
  25. IAB/PwC Internet Advertising Revenue Report 2024.
- 

**Украинский Константин Викторович**, аспирант, Школа вычислительных социальных наук Европейского университета в Санкт-Петербурге, Санкт-Петербург, Россия  
E-mail: [k.ukrainskii@eu.spb.ru](mailto:k.ukrainskii@eu.spb.ru)  
ORCID: 0009-0005-1688-3957

*Поступила в редакцию 01.12.2025*

UDC 330.47: 519.246.5

DOI 10.5281/zenodo.18048614

**UKRAINSKII Konstantin**<sup>1</sup>

<sup>1</sup> School of Computational Social Sciences of the European University in St. Petersburg, Gagarinskaya str., 6/1, letter A, St. Petersburg, Russia, 191187

## **NEURAL HAWKES PROCESSES FOR INTENSITY PARAMETER ESTIMATION AND OPTIMIZATION OF MARKETING STRATEGIES UNDER LIMITED DATA ACCESS**

The article is dedicated to the development of an interpretable neural model based on Hawkes processes with attention mechanisms for analyzing the dynamics of music track listens on digital platforms. The work substantiates the relevance of assessing the effectiveness of marketing strategies under conditions of limited data access, caused by regulatory restrictions. Current issues in data collection and processing are examined, including event timestamps and categorical variables such as listening source, age group, and time of day. Empirical studies of user behavior are explored, identifying groups of "specialists," "generalists," and heavy users, emphasizing the need to account for consumption diversity and self-excitation effects.

The econometric model of repeated discrete choice is analyzed in detail, presenting demand dynamics as an aggregated event counter, considering the proportion of consumers revising their choices and endogenous factors. Limitations of traditional approaches are characterized, including Pareto/NBD models and classical Hawkes processes, related to event independence and exponential parameter growth. The advantages of the proposed model are highlighted, integrating vector representations and transformer architecture to capture self-excitation and inhibition.

A key aspect of the study is the methodology, including data preprocessing, intensity modeling with extension to marked events, division into training and validation samples, quality assessment through residual analysis and regression metrics. Mechanisms for separate training of event categories and evaluation of independent variable importance via attention weights are presented. The article includes analysis of impact kernels, cumulative influence of event types, and optimal time lags. Planned results are reflected in tables and diagrams, demonstrating asymmetry of influences and practical value for campaign optimization.

Conclusions are drawn about the model's ability to overcome data limitations, ensuring accurate forecasting of behavior dynamics and scenario simulation. Continuous intensity analysis will reveal loyalty and saturation effects, contributing to increased returns on marketing investments in the digital economy. Diagnosis of response temporal structure will determine opportunities for adaptive strategies, clearly delineating the model's impact boundaries on targeting optimization.

**Key words:** *Hawkes neural processes; attention mechanisms; user behavior modeling; digital platforms; intensity assessment; marketing strategies; temporal dynamics; vector representations; targeting optimization.*

### **References**

1. Xu L., Duan J. A., Whinston A. Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion. *Management Science. Institute for Operations Research and the Management Sciences (INFORMS)*, 2014. Т. 60, № 6. С. 1392–1412.
2. Anderson A. и др. Algorithmic Effects on the Diversity of Consumption on Spotify // *Proceedings of The Web Conference 2020*. Taipei Taiwan: ACM, 2020. С. 55–65.

3. Datta H., Knox G., Bronnenberg B. J. Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery. *Marketing Science*. 2018. Т. 37, № 1. С. 5–21.
  4. Morris J. W., Powers D. Control, curation and musical experience in streaming music services: 2. *Creative Industries Journal*. 2015. Т. 8, № 2. С. 106–122.
  5. Anas A. Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological*. 1983. Т. 17, № 1. С. 13–23.
  6. Ben-Akiva M. и др. Modeling Methods for Discrete Choice Analysis. *Marketing Letters*. 1997. Т. 8, № 3. С. 273–286.
  7. Leake C. Discrete Choice Theory of Product Differentiation. *Journal of the Operational Research Society*. 1995. Т. 46, № 4. С. 543–543.
  8. Cowan R., Cowan W., Peter Swann G. M. Waves in consumption with interdependence among consumers. *Canadian J of Economics*. 2004. Т. 37, № 1. С. 149–177.
  9. Fader P. S., Hardie B. G. S., Lee K. L. “Counting Your Customers” the Easy Way: An Alternative to the Pareto/NBD Model. *Marketing Science. Institute for Operations Research and the Management Sciences (INFORMS)*, 2005. Т. 24, № 2. С. 275–284.
  10. Hawkes A. G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika. Oxford University Press (OUP)*, 1971. Т. 58, № 1. С. 83–90.
  11. Isham V., Westcott M. A self-correcting point process. *Stochastic Processes and their Applications. Elsevier BV*, 1979. Т. 8, № 3. С. 335–347.
  12. Eichler M., Dahlhaus R., Dueck J. Graphical Modeling for Multivariate Hawkes Processes with Nonparametric Link Functions. arXiv, 2016.
  13. Du N. и др. Recurrent Marked Temporal Point Processes: Embedding Event History to Vector // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, 2016. С. 1555–1564.
  14. Mei H., Eisner J. The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process. arXiv, 2016.
  15. Meng Z. и др. Interpretable Transformer Hawkes Processes: Unveiling Complex Interactions in Social Networks. arXiv, 2024.
  16. Zhao Y., Engelhard M. Balancing Interpretability and Flexibility in Modeling Diagnostic Trajectories with an Embedded Neural Hawkes Process Model. arXiv, 2025.
  17. Zuo S. и др. Transformer Hawkes Process. arXiv, 2020.
  18. Karpukhin I., Dereka S., Kolesnikov S. EXACT: How to train your accuracy. *Pattern Recognition Letters*. 2024. Т. 185. С. 23–30.
  19. Meng Z. и др. TransFeat-TPP: An Interpretable Deep Covariate Temporal Point Processes // Frontiers in Artificial Intelligence and Applications / под ред. Endriss U. и др. IOS Press, 2024.
  20. Kirchner M. An estimation procedure for the Hawkes process. arXiv, 2015.
  21. Zuo S. и др. Transformer Hawkes Process. arXiv, 2020.
  22. Ogata Y. Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Association. JSTOR*, 1988. Т. 83, № 401. Р. 9.
  23. Vaswani A. и др. Attention Is All You Need: arXiv:1706.03762. arXiv, 2023.
  24. Deutsch I., Ross G. J. Estimating Product Cannibalisation in Wholesale using Multivariate Hawkes Processes with Inhibition. arXiv, 2022.
  25. IAB/PwC Internet Advertising Revenue Report 2024.
-

**Ukrainskii Konstantin**, Postgraduate student, School of Computational Social Sciences of the European University in St. Petersburg, Saint-Petersburg, Russia

E-mail: [k.ukrainskii@eu.spb.ru](mailto:k.ukrainskii@eu.spb.ru)

ORCID: 0009-0005-1688-3957

*Received 01.12.2025*