



DOI: 10.22363/2312-8143-2021-22-1-16-22
УДК 004.942

Научная статья / Research article

Анализ применимости методов прогнозирования в системе выбора персонализированных предложений путем аналитического моделирования

Ю.С. Федоренко

Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет),
Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1
E-mail: fedyura11235@mail.ru

История статьи

Поступила в редакцию: 12 февраля 2021 г.
Доработана: 24 апреля 2021 г.
Принята к публикации: 11 мая 2021 г.

Ключевые слова: рекомендательная система, аналитическая модель, система массового обслуживания, время пребывания заявки, функция распределения, переопределение

Аннотация. Актуальность исследования обоснована частым возникновением необходимости решения задач выбора персонализированных предложений в информационных системах и множеством возможных методов машинного обучения, среди которых необходимо выбрать наиболее подходящий. Цель данного исследования — моделирование системы выбора персонализированных предложений как системы массового обслуживания для оценки затрат на оборудование при использовании каждого из методов, необходимых для обслуживания требуемой доли заявок за заданный лимит времени. При этом решается задача оценки минимального количества обслуживающих устройств (серверов подбора), необходимых для обеспечения работы системы на заданном уровне. В работе показано, что систему можно описать многоканальной системой массового обслуживания без отказов. Произведен расчет функции распределения времени пребывания заявки в системе (время обслуживания плюс время ожидания в очереди), так как в литературе для подобных систем описана только функция распределения времени ожидания в очереди. Приведены преобразования выражения вероятности ожидания в системе, решающие проблему переопределения при программной реализации вычисления данного выражения. В заключительной части в качестве примера произведено моделирование системы по заданным параметрам, сделана оценка минимального количества обслуживающих устройств для обеспечения заданного времени ответа системы. По полученным данным можно принять решение о целесообразности применения того или иного метода прогнозирования частоты кликов пользователя или ранжирования.

Для цитирования

Федоренко Ю.С. Анализ применимости методов прогнозирования в системе выбора персонализированных предложений путем аналитического моделирования // Вестник Российского университета дружбы народов. Серия: Инженерные исследования. 2021. Т. 22. № 1. С. 16–22. <http://dx.doi.org/10.22363/2312-8143-2021-22-1-16-22>

© Федоренко Ю.С., 2021



This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by/4.0/>

Applicability analysis of prediction methods in the system for selection personalized offers by analytical modeling

Yuriy S. Fedorenko

Bauman Moscow State Technical University (National Research University of Technology),
5 2-ya Baumanskaya St, bldg 1, Moscow, 105005, Russian Federation
E-mail: fedyura11235@mail.ru

Article history

Received: February 12, 2021

Revised: April 24, 2021

Accepted: May 11, 2021

Abstract. The relevance of the work is justified by the frequent occurrence of the need to solve the problems of choosing personalized offers in information systems and the many possible methods of machine learning, among which it is necessary to choose the most suitable one. The purpose of this study is to simulate a system for selecting personalized offers as a queuing system for estimating equipment costs when using each of the methods necessary to service the required part of requests for a given time limit. This solves the problem of assessing the minimum number of servicing devices (backend servers) required to ensure the operation of the system at a given level. The paper shows that the system can be described by a multichannel queuing system without losses. The distribution function of the spent time of the request in the system (the service time plus the waiting time in the queue) is calculated, since in the literature for such systems only the distribution function of the waiting time in the queue is described. Transformations of the expression for the probability of waiting are given, which solve the overflow problem in the software implementation. In the final part, as an example, the system was modeled according to the given parameters, and the minimum number of servicing devices was estimated to ensure a given system response time. Based on the data obtained, it is possible to make a decision on the advisability of using one or another method for predicting the frequency of user clicks or ranking.

Keywords: recommender system, analytical model, queueing system, mean waiting time, distribution function, overflow

For citation

Fedorenko YuS. Applicability analysis of prediction methods in the system for selection personalized offers by analytical modeling. *RUDN Journal of Engineering Researches*. 2021;22(1):16–22. (In Russ.) <http://dx.doi.org/10.22363/2312-8143-2021-22-1-16-22>

Введение

В современных информационных системах все чаще возникает задача построения систем выбора персонализированных предложений. Это во многом связано с реализацией через интернет ряда товаров и услуг, откуда вытекает применение персонализации предложений для каждого пользователя в таргетированной рекламе и рекомендательных системах. Для решения подобных задач потенциально может быть применен целый спектр методов машинного обучения, таких как деревья решений, глубокие нейросети, метод опорных векторов, а также более специфичные под-

ходы, такие как логистическая регрессия с вручную конструируемыми признаками или специализированными модификациями нейросетей [1]. Однако далеко не все вышеперечисленные методы могут быть применены в конкретных системах с учетом их специфики. Одна из существенных проблем — это ограничение на время ответа системы, которое напрямую зависит от времени получения прогноза.

В качестве примера рассмотрим системы таргетированного показа рекламы. Среднее время ответа сервера по данным сервиса измерения скорости загрузки сайтов <https://webo.in/> составляет 0,1 с у mail.ru и 0,15 с у yandex.ru. За ещё меньшее время должен

быть обработан запрос на подбор рекламы, но будем считать, что это тоже 0,1 с. Предположим, на разные этапы обработки запроса (за исключением получения прогноза) уходит около 0,01 с. Остальное может быть расходувано на получение прогноза частоты кликов (будем считать, что это 0,1 — 0,01 = 0,09 с). Допустим, что в среднем приходится осуществить выбор из 50 баннеров (т.е. получить 50 оценок). Тогда время прогнозирования частоты кликов для каждой пары пользователь–баннер должно составлять не более $0,09/50 = 1,8 \cdot 10^{-3}$ с (параллельно рассчитывать оценки для каждой пары пользователь–баннер нецелесообразно, поскольку потоки сервера обычно заняты обработкой других запросов). Это накладывает ограничения на используемые методы прогнозирования.

В данной работе предложено моделировать систему выбора персонализированных предложений как систему массового обслуживания (СМО) для оценки применимости различных методов машинного обучения с точки зрения временных требований к системе.

1. Постановка задачи

Будем считать, что запросы на показ рекламы (заявки) поступают с одинаковой интенсивностью λ . В таком случае поток заявок можно считать стационарным (вероятность появления определенного числа заявок на заданном интервале зависит только от продолжительности данного интервала), а сами заявки независимыми друг от друга (пользователи открывают страницы в интернете независимо друг от друга). Также будем предполагать, что на временном интервале достаточно малой длительности δt приходит только 1 заявка. Следовательно, в рассматриваемой системе поток заявок является простейшим, а промежуток времени между двумя заявками имеет экспоненциальное распределение с параметром λ .

Таким образом, систему выбора персонализированных предложений можно рассматривать как многоканальную СМО без отказов (заявки не покидают очередь) с количеством обслуживающих устройств m . При этом имеется желательный лимит на суммарное время ожидания и обслуживания T (время пребывания заявки в системе). При функционировании системы требуется,

чтобы доля необработанных за данное время запросов не превышала γ от общего числа запросов. Требуется оценить минимальное количество обслуживающих устройств m (серверов подбора рекламы), необходимых для обеспечения работы системы на заданном уровне.

2. Аналитическая модель рекламной системы как системы массового обслуживания

Описанная выше СМО традиционно обозначается как $M/M/m$, время обслуживания заявки считаем распределенным по экспоненциальному закону с интенсивностью μ (среднее время обслуживания равняется $1/\mu$). Схема такой СМО представлена на рис. 1. Математическая модель данной СМО описана в литературе¹. Воспользуемся данными результатами для анализа системы выбора персонализированных предложений. Будем рассматривать стационарный режим работы системы. Обозначим приведенную плотность потока заявок через $\alpha = \frac{\lambda}{\mu}$. Будем считать, что $\frac{\alpha}{m} < 1$, так как в противном случае рассматривать систему смысла не имеет.

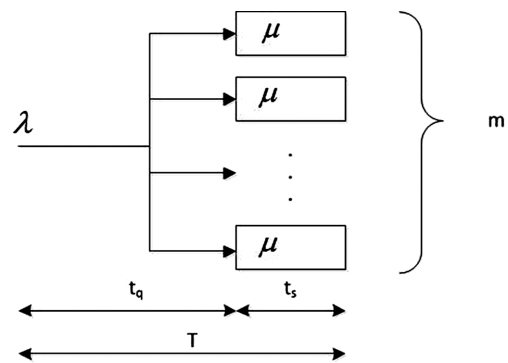


Рис. 1. Схема СМО $M/M/m$

Figure 1. $M/M/m$ QS scheme

Вероятность того, что в системе нет ни одного запроса, равняется:

$$p_0 = \left(\sum_{i=0}^{m-1} \frac{\alpha^i}{i!} + \frac{\alpha^m}{m! (m - \alpha)} \right)^{-1}. \quad (1)$$

¹ Каштанов В.А., Ивченко Г.И., Коваленко И.Н. Теория массового обслуживания: учебное пособие. Изд. 2-е, испр. и доп. М.: Книжный дом ЛИБРОКОМ, 2012. 304 с.

Вероятность ожидания в системе p_w равняется:

$$p_w = \frac{p_0 \cdot \alpha^m}{(m-1)!(m-\alpha)}. \quad (2)$$

Плотность распределения времени ожидания в такой системе равняется [3]:

$$f_q(t) = \begin{cases} (1-p_w) \cdot u_0(t) + p_w \cdot (m \cdot \mu - \lambda) e^{-t(m \cdot \mu - \lambda)}, & t > 0; \\ 0, & t < 0, \end{cases} \quad (3)$$

где $u_0(t)$ — импульсная функция в начале координат (она появляется для нормировки плотности распределения вероятности).

Плотность распределения для времени обслуживания традиционно равняется:

$$f_{q+s}(t) = \begin{cases} \mu \cdot e^{-\mu t}, & t > 0; \\ 0, & t < 0. \end{cases} \quad (4)$$

Требуется найти плотность распределения f_{q+s} .

Плотность распределения случайной величины, равной сумме двух независимых случайных величин, рассчитывается по правилу свертки²:

$$f_{q+s}(t) = \int_{-\infty}^{+\infty} f_q(u) \cdot f_s(t-u) du. \quad (5)$$

Возьмем данный интеграл для заданных плотностей распределения. Поскольку плотности распределения f_s и f_q имеют ненулевое значение только при положительных значениях аргумента (см. (3) и (4)), в выражении (5) ненулевое значение получается при $u > 0$ и $t - u > 0$. Следовательно, пределы интегрирования по u следует изменить от 0 до t . Тогда получаем:

$$\begin{aligned} f_{q+s}(t) &= \int_0^t \left((1-p_w) \cdot u_0(u) + \right. \\ &+ \left. p_w \cdot (m \cdot \mu - \lambda) e^{-u(m \cdot \mu - \lambda)} \right) \cdot \mu \cdot e^{-\mu(t-u)} du = \\ &= \int_0^t ((1-p_w) \cdot u_0(u) \cdot \mu \cdot e^{-\mu(t-u)} du + \\ &+ \int_0^t p_w \cdot (m \cdot \mu - \lambda) e^{-u(m \cdot \mu - \lambda)} \cdot \mu \cdot e^{-\mu(t-u)} du. \end{aligned}$$

Рассмотрим каждый интеграл в отдельности.

$$\begin{aligned} &\int_0^t ((1-p_w) \cdot u_0(u) \cdot \mu \cdot e^{-\mu(t-u)} du = \\ &= \mu \cdot (1-p_w) \int_0^t u_0(u) e^{-\mu(t-u)} du = \\ &= \mu \cdot (1-p_w) \int_0^t (e^{-\mu \cdot u} \cdot u_0(t-u) du). \end{aligned}$$

Последнее равенство верно в силу симметричности свертки функций. Согласно фильтрующему свойству дельта функции [3], $\int_a^b f(t) u_0(t_0 - t) dt = f(t_0)$ при $a < t_0 < b$. Следовательно:

$$\begin{aligned} &\mu \cdot (1-p_w) \cdot \int_0^t (e^{-\mu \cdot u} \cdot u_0(t-u) du) = \\ &= \mu \cdot (1-p_w) \cdot e^{-\mu t}, \end{aligned}$$

поскольку $0 < u < t$.

Рассмотрим второе слагаемое:

$$\begin{aligned} &\int_0^t p_w \cdot (m \cdot \mu - \lambda) e^{-u(m \cdot \mu - \lambda)} \cdot \mu \cdot e^{-\mu(t-u)} du = \\ &= p_w \cdot (m \cdot \mu - \lambda) \cdot \mu \int_0^t e^{-(m \cdot \mu - \lambda)u - \mu(t-u)} du = \\ &= p_w \cdot (m \cdot \mu - \lambda) \cdot \mu e^{-\mu t} \int_0^t e^{-u((m-1) \cdot \mu - \lambda)} du = \\ &= \frac{p_w \cdot (m \cdot \mu - \lambda) \cdot \mu e^{-\mu t}}{(m-1) \cdot \mu - \lambda} \cdot (1 - e^{-t((m-1) \cdot \mu - \lambda)}). \end{aligned}$$

Складывая оба полученных слагаемых, получаем, что плотность распределения времени пребывания заявки в системе равняется:

$$\begin{aligned} f_{q+s}(t) &= \mu \cdot (1-p_w) \cdot e^{-\mu t} + \\ &+ \frac{p_w \cdot (m \cdot \mu - \lambda) \cdot \mu e^{-\mu t}}{(m-1) \cdot \mu - \lambda} \cdot (1 - e^{-t((m-1) \cdot \mu - \lambda)}) = \\ &= \mu \cdot e^{-\mu t} \left(1 + \frac{p_w \cdot \mu}{(m-1) \cdot \mu - \lambda} \right) - \\ &- \frac{p_w \cdot (m \cdot \mu - \lambda) \cdot \mu}{(m-1) \cdot \mu - \lambda} e^{-t(m \cdot \mu - \lambda)}. \end{aligned}$$

² Печинкин А.В., Тескин О.И., Цветкова Г.М. Теория вероятностей: учеб. для вузов. 3-е изд., испр. / под ред. Зарубина В.С., Крищенко А.П. М.: Изд-во МГТУ им. Н.Э. Баумана, 2004. 456 с.

Итого получаем, что плотность распределения времени ожидания и обслуживания равняется:

$$f_{q+s}(t) = \begin{cases} \mu \cdot e^{-\mu t} \left(1 + \frac{p_w \cdot \mu}{(m-1) \cdot \mu - \lambda}\right) - \frac{p_w \cdot (m \cdot \mu - \lambda) \cdot \mu}{(m-1) \cdot \mu - \lambda} e^{-t(m \cdot \mu - \lambda)}, & t > 0 \\ 0, & t < 0 \end{cases}$$

Нетрудно проверить, что интеграл от данной функции по всем значениям t равняется 1. Для получения функции распределения данной случайной величины необходимо проинтегрировать ее плотность распределения:

$$F_{q+s}(t) = \int_{-\infty}^t f_{q+s}(u) du.$$

Поскольку плотность распределения f_{q+s} имеет ненулевое значение при положительном значении аргумента, получаем, что пределы интегрирования следует изменить от 0 до t . Кроме того, для краткости записи, обозначим $\frac{p_w \cdot \mu}{(m-1) \cdot \mu - \lambda}$ через c . Тогда можно записать

$$\begin{aligned} F_{q+s}(t) &= \int_0^t f_{q+s}(u) du = \int_0^t (\mu \cdot e^{-\mu \cdot u} \cdot (1 + c) - \\ &- c \cdot (m \cdot \mu - \lambda) \cdot e^{-u(m \cdot \mu - \lambda)}) du = \\ &= (1 + c) \int_0^t \mu \cdot e^{-\mu \cdot u} du - \\ &- c \cdot \int_0^t (m \cdot \mu - \lambda) \cdot e^{-u(m \cdot \mu - \lambda)} du = \\ &= 1 - e^{-\mu \cdot t} \cdot (1 + c) + c \cdot e^{-(m \cdot \mu - \lambda) \cdot t}. \end{aligned}$$

Согласно требованиям модели, за заданный лимит времени T должно быть обработано $1 - \gamma$ запросов. Отсюда можно записать

$$F_{q+s}(T) > 1 - \gamma \rightarrow 1 - \gamma < 1 - e^{-\mu \cdot T} \cdot (1 + c) + c \cdot e^{-(m \cdot \mu - \lambda) \cdot T}.$$

После небольших преобразований получаем

$$e^{-T \cdot \mu} \cdot (1 + c) - c \cdot e^{-T(m \cdot \mu - \lambda)} < \gamma.$$

Соответственно, требуется определить, при каком минимальном значении m выполняется данное неравенство. Ясно, что получить аналитическое выражение для m через известные параметры не удастся, однако на основании полученных выражений можно составить график доли запросов γ , которые не успели обработаться за заданное время, от количества серверов m . Затем с его помощью можно оценить

необходимое число серверов для разной сложности модели прогнозирования.

Однако следует отметить, что при моделировании системы выбора персонализированных предложений параметры m и α принимают большие значения в несколько сотен, а то и тысяч (так как имеется минимум несколько десятков серверов, каждый из которых обрабатывает по паре десятков запросов). В результате вычисление α^m и $m!$ приводит к переполнению. Например, в языке python при использовании экспоненциальной формы представления чисел максимальное значение порядка степени составляет 307—308, т.е. максимально возможное представимое число имеет порядок $10^{307-308}$. Однако, к примеру, даже $200^{200} = 10^{\log 200 \cdot 200} \approx 10^{460}$, т.е. наступает переполнение. Решение этой проблемы с помощью специализированных библиотек для работы с большими числами также не есть хороший путь, поскольку будут возрастать погрешности, тем более при реальном моделировании системы выбора персонализированных предложений числа получаются ещё значительно больше и можно все равно получить значение, большее, чем максимально допустимое в библиотеке. По этой причине предлагается провести определенное преобразование формул (1) и (2). Изначально имеем:

$$p_w = \frac{\alpha^m}{(m-1)!(m-\alpha)} \cdot p_0 = \frac{\alpha^m}{\left(\sum_{i=0}^m \frac{\alpha^i}{i!} + \frac{\alpha^{m+1}}{m!(m-\alpha)}\right)}.$$

Разделив числитель и знаменатель на α^m , получим

$$p_w = \frac{1}{\left(\sum_{i=0}^m \frac{1}{i! \cdot \alpha^{m-i}} + \frac{\alpha}{m!(m-\alpha)}\right)}.$$

Домножив числитель и знаменатель на $m!$, получим

$$p_w = \frac{m!}{\left(\sum_{i=0}^m \frac{m!}{i! \cdot \alpha^{m-i}} + \frac{\alpha}{(m-\alpha)}\right)}. \quad (6)$$

Слагаемое в сумме можно вычислить следующим образом:

$$\begin{aligned} \frac{m!}{i! \cdot \alpha^{m-i}} &= \frac{m \cdot (m-1) \cdot \dots \cdot (i+1)}{\alpha^{m-i}} = \\ &= \frac{\overbrace{m \cdot m-1 \cdot \dots \cdot i+1}^{m-i \text{ множителей}}}{\alpha}. \end{aligned} \quad (7)$$

При программной реализации модели такое выражение легко вычисляется в обычном цикле. Для суммирования достаточно написать еще один цикл. Таким образом, значение вероятности ожидания в системе p_w предлагается вычислять по формулам (6) и (7).

3. Пример моделирования рекламной системы как системы массового обслуживания

В качестве примера оценим количество обслуживающих устройств (серверов подбора рекламы), необходимых для удовлетворительной работы системы. Будем рассматривать 3 модели прогнозирования [4]:

1. Логистическая регрессия с хешированием комбинаций признаков (lr);
2. Нейросеть, перебирающая пары признаков (nn2) [3];
3. Нейросеть, перебирающая тройки признаков (nn3) [3].

Зададим следующие значения параметров:

- $\lambda = 40000$ запросов/с;
- $t_{other} = 0,01$ с;
- $t_{lr} = 0,0002 \times 50 = 0,01$ с;
- $t_{nn2} = 0,00022 \times 50 = 0,011$ с;
- $t_{nn3} = 0,0003 \times 50 = 0,015$ с;
- $\mu = 1 / (t_{other} + (t_{lr} | t_{nn2} | t_{nn3}))$ 1/с;
- $T = 0,1$ с.

Параметр λ был рассчитан на основании данных из открытых источников [4] по аудитории крупнейших сайтов в РФ и по среднему количеству просматриваемых баннеров (для крупных проектов в интернете с миллионами пользователей эта цифра может достигать и сотен тысяч запросов в секунду).

Время, затрачиваемое на работу моделей t_{lr} , t_{nn2} , t_{nn3} , складывается из времени прогнозирования одного примера, умноженного на примерное число баннеров, из которых производится выбор (при моделировании будем считать, что это 50 баннеров). Время обслуживания складывается из времени, затрачиваемого на получение оценок для баннеров и прочие служебные операции t_{other} . Интенсивность обслуживания μ обратно пропорциональна среднему времени обслуживания (обработки запроса). Будем считать, что для удовлетворительной работы системы выбора персонализированных пред-

ложений необходимо, чтобы не более 1 % запросов обрабатывались более времени $T = 0,1$ с.

При моделировании изменялось количество обслуживающих устройств m и измерялось значение γ — доли запросов, которые обрабатывались более времени T . Результаты моделирования приведены на рис. 2.

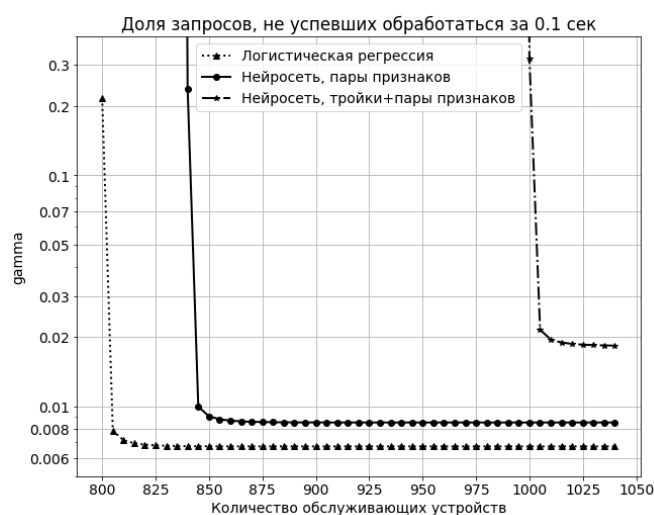


Рис. 2. Результаты моделирования системы выбора персонализированных предложений как СМО

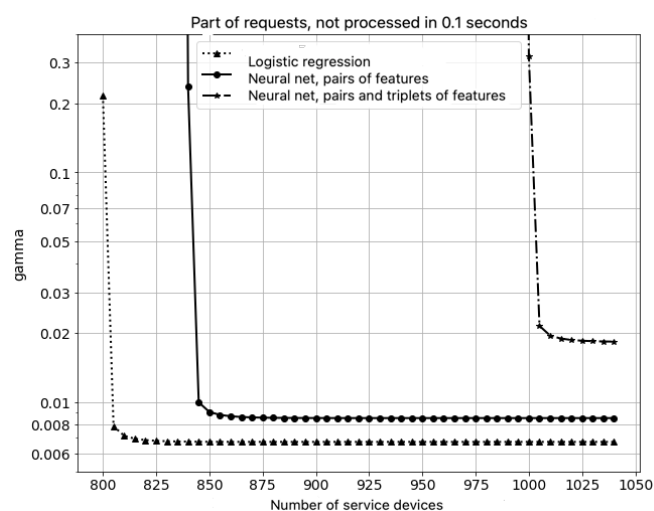


Figure 2. Results of modeling the system for selecting personalized offers as a QS

По результатам моделирования видно, что 810—815 обслуживающих устройств (около 82 серверов, обрабатывающих по 10 запросов одновременно) доста-

точно для обеспечения работы системы при использовании модели логистической регрессии (доля запросов, не успевших обработаться за 0,1 с, составляет около 0,7 %). При использовании нейросети с перебором пар признаков требуется 850—855 обслуживающих устройств (доля запросов, не успевших обработаться за 0,1 с, составляет около 0,9 %). При использовании модели нейросети с перебором троек признаков для стабильной работы требуется около 1020 обслуживающих устройств, но при этом доля запросов, не успевших обработаться за 0,1 с, все равно превышает лимит, составляя около 1,9 %). С учетом получаемой прибыли от работы каждой модели можно сделать вывод о целесообразности ее применения.

Заключение

Таким образом, предложенная в работе аналитическая модель системы выбора персонализированных предложений как СМО позволяет принимать решения относительно целесообразности использования того или иного алгоритма прогнозирования в системе. По результатам моделирования формируется картина затрат на оборудование при использовании каждой из моделей, необходимых для достижения целевого времени обработки заявок в системе (чтобы требуемая доля заявок была обслужена за заданный лимит времени). С учетом данных о прибыли системы при работе с каждой из моделей можно принять решение о том, какую модель стоит использовать на практике.

Сведения об авторе

Федоренко Юрий Сергеевич, соискатель, кафедра систем обработки информации и управления, факультет информатики и системы управления, МГТУ им. Н.Э. Баумана; eLIBRARY SPIN-код: 1755-4017, Scopus Id: 57192874789; e-mail: fedyura11235@mail.ru

About the author

Yuriy S. Fedorenko, Degree Seeker at the Department of Informatics and Control Systems, Faculty of Computer Science and Management Systems, BMSTU; eLIBRARY SPIN-code: 1755-4017, Scopus Id: 57192874789; e-mail: fedyura11235@mail.ru

Список литературы

1. Fedorenko Y.S., Chernenkiy V.M., Gapanyuk Y.E. The Neural Network for Online Learning Task Without Manual Feature Extraction // *Advances in Neural Networks*. Cham: Springer, 2019. Vol. 11554. P. 67—77. http://dx.doi.org/10.1007/978-3-030-22796-8_8
2. Тихонов В.И., Миронов М.А. Марковские процессы. М.: Сов. Радио, 1977. 485 с.
3. Федоренко Ю.С. Проектирование быстрой программной реализации специализированной нейросетевой архитектуры с разреженными связями // Программные продукты и системы. 2019. № 4. С. 639—649. <http://dx.doi.org/10.15827/0236-235X.128.639-649>
4. Xinran H., Junfeng P., Ou J., Tianbing X., Bo L., Tao H. et al. Practical Lessons from Predicting Clicks on Ads at Facebook // *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*. New York, USA: ACM. 2014. P. 1—9. <http://dx.doi.org/10.1145/2648584.2648589>

References

1. Fedorenko YS, Chernenkiy VM, Gapanyuk YE. The Neural Network for Online Learning Task Without Manual Feature Extraction. *Advances in Neural Networks*. 2019; 11554:67—77. http://dx.doi.org/10.1007/978-3-030-22796-8_8
2. Tihonov VI, Mironov MA. *Markovskie processy [Markov processes]*. Moscow: Sovetskoe radio Publ.; 1977. (In Russ.)
3. Fedorenko YS. The development of fast software implementation of specialized neural network architecture with sparse connections. *Software & Systems*. 2019;32(4):639—649. (In Russ.) <http://dx.doi.org/10.15827/0236-235X.128.639-649>
4. Xinran H, Junfeng P, Ou J, Tianbing X, Bo L, Tao H et al. Practical Lessons from Predicting Clicks on Ads at Facebook. In: *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*. New York, USA: ACM Publ.; 2014. p. 1—9. <http://dx.doi.org/10.1145/2648584.2648589>