

УДК 004.852

DOI:10.25729/ESI.2024.34.2.006

## Оценка качества интерполяции плотности распределения случайной величины при помощи искусственной нейронной сети

Доррер Михаил Георгиевич

Сибирский государственный университет науки и технологий им. М.Ф. Решетнева,  
Россия, Красноярск, [mdorrer@mail.ru](mailto:mdorrer@mail.ru)

**Аннотация.** В работе описаны создание и оценка работоспособности нейросетевой модели функции плотности распределения вероятности случайной величины, заданной набором измерений случайной величины при отсутствии стадии идентификации закона распределения. Потребность в решении данной задачи вызывается ограничениями, вносимыми в точность расчета функции плотности распределения вероятности случайной величины как таблично-гистограммным методом, так и в случае применения подходов к идентификации закона распределения. Задача решалась на языке Python с использованием нейросетевой библиотеки TensorFlow путем создания нейросетевой модели на базе класса Sequential с полносвязными слоями Dense, обученной на данных численного дифференцирования функции распределения случайной величины. Точность прогноза оценивалась с помощью меры расстояния Кульбака-Лейблера для различных соотношений объема экспериментальных данных и количества интервалов интерполяции на синтетических тестовых данных, сгенерированных для 5 законов распределения – Рэлея, Вэйбулла, гамма, экспоненциального и нормального (гауссовского). Для оценки прогностической способности подхода при тестировании интерполятора использовались отсчеты случайной величины, сдвинутые по отношению к используемым при обучении. Предложенное решение показало значительно более высокую точность расчета значений плотности распределения случайной величины по сравнению с гистограммным методом. Разработанный подход будет внедрен в моделирующую часть цифрового двойника бизнес-процесса, основанного на математическом аппарате стохастических GERT-сетей.

**Ключевые слова:** стохастические модели, плотность распределения случайной величины, нейронные сети, цифровые модели процессов

**Цитирование:** Доррер М.Г. Оценка качества интерполяции плотности распределения случайной величины при помощи искусственной нейронной сети / М.Г. Доррер // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 59-68. – DOI:10.25729/ESI.2024.34.2.006.

**Введение.** В настоящее время задачи управления организационными системами постоянно усложняются [1]. Отдельные части организационных систем становятся все более интеллектуальными, число и сложность выполняемых ими функций возрастают. Разные части организационной системы функционируют параллельно и в процессе решения задач постоянно обмениваются между собой информацией. Ведется обмен не только данными, но и во все большей мере звуковой и видеоинформацией, часто в реальном времени. Примерами таких систем являются системы управления бизнес-процессами организации, обеспечивающие взаимодействие пользователей как внутри организации, так и с внешними акторами, и включающие в себя интеллектуальные системы – боты, ассистенты, системы роботизации бизнес-процесса и т.д.

Предлагаемый в данной работе результат был получен в ходе решения задач по созданию цифрового двойника бизнес-процесса на базе системы управления бизнес-процессами [2]. В качестве математического аппарата для моделей аддитивных параметров бизнес-процесса (время выполнения, стоимость, ресурсоемкость) в цифровом двойнике используется модель стохастических GERT-сетей. Для использования этого математического аппарата необходимо было получить решение задачи нахождения непрерывных плотностей распределения данных параметров при прохождении стохастических GERT-сетей с использованием топологического уравнения для оценки вероятностно-временных и вероятностно-ресурсных характеристик организационных систем.

При этом разнородность организационно-технической системы требовала сопряжения нескольких моделей бизнес-процессов, как на одном уровне, так и иерархически. При этом должно было обеспечиваться информационное единство среды, на которой строятся сложные иерархические модели, моделируемые в цифровом двойнике. На ранних этапах реализации это обеспечивалось путем представления распределений времени выполнения отдельных операций (эквивалентных операций, GERT-сетей) их характеристическими функциями в табличном представлении. Такой подход применяется, когда исходная или промежуточная информация получена в виде плотности или функции распределения. Однако применение табличного представления накладывает довольно существенное ограничение на алгоритмы расчета, привязывая их к единой, фиксированной для всей модели интерполяционной сетке.

В данной работе будет рассмотрено решение задачи представления характеристических функций аддитивных параметров операций модели бизнес-процесса в виде искусственной нейронной сети, включая сравнительный анализ данного подхода с базовым, таблично-гистограммным методом расчета функции плотности распределения.

В разделе 1 представлен краткий обзор публикаций, затрагивающих тему цифрового моделирования бизнес-процессов, в разделе 2 – описание постановки задачи и использованных экспериментальных материалов, в разделе 3 показаны применяемые методы решения задачи интерполяции функции плотности распределения вероятности, в разделе 4 описана интерпретация полученных результатов.

**1. Связанные работы.** Работы по моделированию бизнес-процессов с помощью математических моделей в настоящее время разрабатываются достаточно активно, в публикациях описаны несколько частных решений данной задачи. Так, в работе [3] предложен подход к формальному описанию бизнес-процессов с использованием ряда математических ограничений (которые определяют границы осуществимости бизнес-процесса) и набора целевых функций (которые состоят из различных целей для проектирования бизнес-процесса). Аналогичный подход представлен в работе [4], авторы которой описывают математическую модель, содержащую основные компоненты общего бизнес-процесса. В работе [5] предлагается использовать формальные математические обозначения как способ введения бизнес-правил и проверки логической согласованности схематических моделей. В работе [6] предложено теоретико-множественное формальное описание бизнес-процессов, которое, например, ставит задачи разработки алгоритмов верификации бизнес-процесса [7]. Авторы работы [8] предлагают использовать для анализа бизнес-процессов их описание в форме тензоров специального вида и определенных над ними операций. Первоначально этот подход был разработан для решения более общей задачи анализа процессов в сложных сетевых системах [9].

Для решения ряда частных задач по анализу бизнес-процессов используются стохастические GERT-сети (Graphical Evaluation and Review Technique) [10 - 13]. GERT-сеть позволяет моделировать поток работ в виде ориентированного графа, ребрами которого являются задачи бизнес-процесса [14, 15].

Данная работа является продолжением цикла публикаций, посвященных реализации цифрового двойника бизнес-процесса – ЦДБП (Business Process Digital Twin, BPDT) [16], [17], [18], [2].

Актуальность темы обусловлена проблемами в точности прогноза вероятностных характеристик бизнес-процессов, выявленных в построении функции плотности распределения случайной величины, заданной отчетами случайного процесса, таблично-гистограммным методом [19] и путем идентификации законов распределения [20].

**2. Постановка задачи и экспериментальные материалы.** Оценка методов расчета плотности распределения случайной величины выполняется с использованием данных, сгенерированных для 5 распределений, для которых известны аналитические выражения плотности распределения. Работа базируется на методах, предложенных в статье [21] с расширением набора моделируемых законов распределения, а также содержит полностью оригинальный блок создания программного объекта – интерполятора функции плотности распределения случайной величины на базе многослойной полносвязной нейронной сети. Используемый набор законов распределения включает закон Релея, гамма-распределение, распределение Вейбулла, экспоненциальное распределение и нормальное (гауссово) распределение.

Исходные данные в виде массивов измерений случайной величины для 5 вышеперечисленных законов распределения генерируются с помощью кода Python, показанного в листинге 1.

**Листинг 1.** Генерация исходных данных для набора законов распределения

```
# Генерируем плотности распределения в соответствии с 5 законами
pdf_function = {
    'rrand': partial(rel_pdf, 1),
    'grand': partial(gam_pdf, 0.5, 0.5),
    'wrand': partial(weib_pdf, 1, 5),
    'exprand': partial(exp_pdf, 1.5),
    'gaussrand': partial(gauss_pdf, 1.0, 1.0)
}
```

Пример функции **gauss\_pdf**, генерирующей синтетические данные отсчетов случайной величины для нормального распределения, приведен в листинге 2.

**Листинг 2.** Функция генерации исходных данных для нормального закона распределения

```
def gauss_pdf(mu: float, sigma: float, X: list) -> pandas.DataFrame:
    """
```

```
    Вычисляет кривую гауссовской плотности распределения вероятности по известной
    формуле
    :param n: количество отсчетов
    :param mu: матожидание
    :param sigma: дисперсия
    :param X: координаты по оси абсцисс
    :return: pandas.DataFrame
    """
```

```
    pdf_y = [] # Координаты по оси ординат
    for x in X:
        pdf_y.append( (1 / np.sqrt(2 * np.pi * sigma)) * np.exp(-(x - mu)**2 / (2 *
        sigma)))
    return pd.DataFrame({'x': X, 'y': pdf_y})
```

Далее выполним сравнение точности расчета функции плотности распределения случайной величины для вышеперечисленных 5 законов между аналитическими значениями и значениями, полученными следующими методами:

- путем интерполяции таблично-гистограммным методом;
- путем интерполяции полносвязной нейронной сетью, обученной на табличных значениях функции плотности распределения случайной величины, рассчитанной как первая производная от функции распределения, которая, в свою очередь, представляет собой вероятность обнаружить значение случайной величины меньше либо равное заданному [22].

**3. Методы решения задачи.** Сравнение точности исследуемых методов производилось с помощью меры расстояния Кульбака-Лейблера от априорной функции плотности распределения  $P$  до получаемой оценки  $Q$ , рассчитываемой как  $KL = \mathbb{E} \left[ P \cdot \ln \left( \frac{P}{Q} \right) \right]$ .

Далее производится расчет обучающей выборки для нейросетевого интерполятора функции плотности распределения случайной величины. При этом используется следующий алгоритм [22] ( $N$  – количество доступных отсчетов случайной величины):

- 1) сортировка значений отсчетов случайной величины  $X$  по возрастанию, получение упорядоченного набора значений  $\hat{X}$ ;
- 2) сопоставление с каждым значением в  $\hat{X}$  его порядкового номера  $i$ , начиная с нуля, тогда с точностью до множителя  $i(\hat{X}_i)$  представляет собой оценку функции распределения случайной величины;
- 3) построение равномерной шкалы из  $k + 2$  значений на интервале от  $X_{min}$  до  $X_{max}$  где  $k$  – желаемое число узлов сетки, в которых рассчитывается функция плотности распределения ( $k < N$ );
- 4) интерполяция значений номеров переменных из шкалы упорядоченного массива значений переменной в равномерную шкалу, полученную в п. 3;
- 5) численное дифференцирование интерполированной функции по соседним точкам (для чего сетка и состоит из  $k + 2$  значений) с делением каждого значения на  $N$  – в результате получается искомая оценка плотности вероятности.

В результате выполнения данного алгоритма формируется переменная  $\mathbf{d}$  – таблица значений плотности распределения случайной величины. В реализованном в рамках настоящего исследования примере она имеет вид DataFrame из библиотеки Pandas, содержащего два столбца вещественных чисел – «x» – значения случайной величины и «y» – значения плотности распределения для данного значения случайной величины.

Далее на основе полученного датафрейма  $\mathbf{d}$  выполняется формирование и обучение модели нейронной сети, позволяющей получить значение функции плотности распределения не только в узлах сетки, содержащейся в обучающей выборке, но и на всей области определения. Код Python для решения задачи приведен в листинге 3.

**Листинг. 3.** Код создания и обучения нейронной сети – интерполятора

```
def generate_neural(k: int, rnd_list: list) -> Sequential:
    """
    Получает кривую плотности распределения вероятности
    :param k: количество интервалов разбиения гистограммы
    :param rnd_list: случайный процесс
    :return: pandas.DataFrame
    """
    pdf_x = [] # Координаты по оси абсцисс
    pdf_y = [] # Координаты по оси ординат
    n = len(rnd_list) # количество элементов в рассматриваемой выборке
    h = (max(rnd_list) - min(rnd_list)) / k # ширина одного интервала
    a = min(rnd_list) # минимальное значение в рассматриваемой выборке
    rnd_list = sorted(rnd_list) # сортируем значения
    j = 0 # индекс значения левой границы интервала
    for i in range(0, k): # проход по интервалам
        count = 0
        while j < n and (a + i * h) <= rnd_list[j] < (a + (i * h) + h): #
            подсчитываем количество значений в k-м интервале
            count = count + 1
            j += 1
        pdf_x.append(a + i * h + h / 2) # координата по оси абсцисс полученной
        # кривой плотности распределения
        pdf_y.append(count / n) # вероятность
```

```

pdf_y.append(count / (n * h)) # координата по оси ординат полученной
кривой плотности распределения
# вероятности
d = {'x': pdf_x, 'y': pdf_y}
x = np.array(d['x'])
y = np.array(d['y'])
Regressor = Sequential()
Regressor.add(Dense(32, input_dim=1, activation='tanh'))
Regressor.add(Dense(16, activation='tanh'))
Regressor.add(Dense(1, activation='sigmoid'))
# Определить функцию ошибки (Loss function) и метод оптимизации (optimizer)
для обучения модели.
Regressor.compile(loss='mse', optimizer='adam', metrics=['accuracy'])
# Обучим модель на данных плотности распределения, используя метод fit() в
Keras.
Regressor.fit(x, y, epochs=100, verbose = 0)
# Передаем результаты для оценки
return Regressor

```

Полученный объект **Regressor** передается в качестве выхода функции формирования объекта-интерполятора и может быть далее использован как встроенный объект, позволяющий получить значение функции плотности распределения случайной величины для любой точки области определения данной функции.

Его применение для построения функции плотности распределения приведено в листинге 4.

**Листинг 4.** Код формирования функции плотности распределения с помощью нейросети – интерполятора Regressor

```

def pdf_neural(k: int, rnd_list: list, Regressor: Sequential) -> pandas.DataFrame:
    """
    Получает кривую плотности распределения вероятности
    :param k: количечиво интервалов разбиения гистограммы
    :param rnd_list: случайный процесс
    :return: pandas.DataFrame
    """
    pdf_x = [] # Координаты по оси абсцисс
    pdf_y = [] # Координаты по оси ординат
    n = len(rnd_list) # количество элементов в рассматриваемой выборке
    h = (max(rnd_list) - min(rnd_list)) / k # ширина одного интервала
    a = min(rnd_list) # минимальное значение в рассматриваемой выборке
    rnd_list = sorted(rnd_list) # сортируем значения
    j = 0 # индекс значения левой границы интервала
    for i in range(0, k): # проход по интервалам
        count = 0
        while j < n and (a + i * h) <= rnd_list[j] < (a + (i * h) + h): #
            подсчитываем количество значений в k-м интервале
            count = count + 1
            j += 1
        pdf_x.append(a + i * h + h / 2) # координата по оси абсцисс полученной
        кривой плотности распределения
    pdf_y = Regressor.predict(pdf_x, verbose = 0)
    d = {'x': pdf_x, 'y': pdf_y}
    x = np.array(d['x'])
    y = np.array(d['y'])
    print(x.shape)
    return d

```

Следует отметить, что для оценки прогностической способности модели ее обучение строится на одной выборке точек, а тестирование – на другой, специально сгенерированной для этой цели. Фрагмент кода, генерирующий отсчеты случайной величины для 5 анализируемых законов распределения в базовом и тестовом варианте, приведён в листинге 5.

**Листинг 5.** Код создания выборок случайных величин для заданных законов

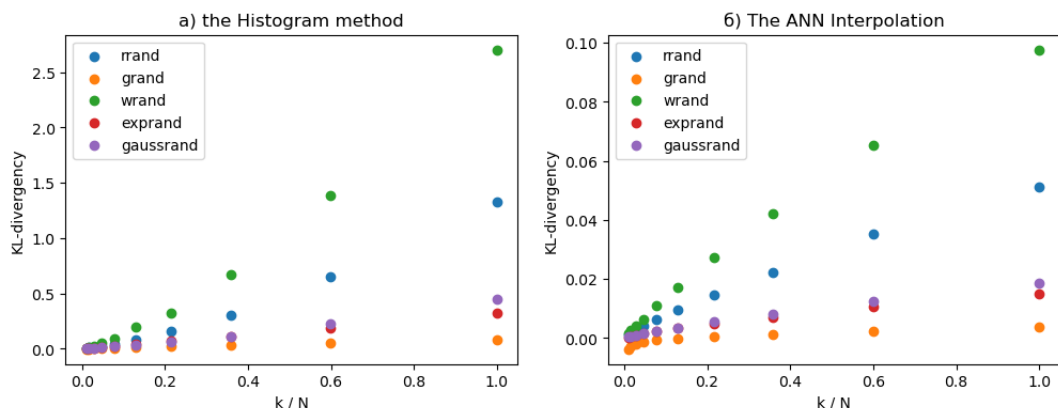
```
# Генерируем случайные серии в соответствии с 5 законами распределения
random_series = {
    'rrand': [rel_rand(N, 1), rel_rand(Ntrain, 1)], # генерируем случайные числа
              с распределением Релея
    'grand': [gam_rand(N, 0.5, 0.5), gam_rand(Ntrain, 0.5, 0.5)], # генерируем
              случайные числа с гамма распределением
    'wrand': [weib_rand(N, 1, 5), weib_rand(Ntrain, 1, 5)], # генерируем случайные
              числа с распределением Вейбулла
    'exprand': [exp_rand(N, 1.5), exp_rand(Ntrain, 1.5)], # генерируем случайные
               числа с экспоненциальным распределением
    'gaussrand': [gauss_rand(N, 1.0, 1.0), gauss_rand(Ntrain, 1.0, 1.0)] #
                 генерируем случайные числа с нормальным распределением
}
```

С учетом этой структуры код, формирующий оценку отклонения рассчитанной аналитически плотности распределения для каждого из 5 анализируемых законов от значений, предсказанных при помощи нейросетевой модели на базе меры Кульбака-Лейблера, выглядит так, как это показано в листинге 6.

**Листинг 6.** Оценка отклонения прогноза плотности распределения

```
metrics_neural = {}
for key, val in random_series.items():
    for k in k_values:
        Regressor = generate_neural(k, val[1][:Ntrain])
        #Regressor = generate_neural(k, val[:N])
        estimated_pdf = pdf_neural(k, val[0][:N], Regressor)
        theoretical_pdf = pdf_function[key](estimated_pdf['x'])
        metrics_neural.setdefault(key, []).append(
            KL_dist(theoretical_pdf['y'], estimated_pdf['y']))
'gaussrand': [gauss_rand(N, 1.0, 1.0), gauss_rand(Ntrain, 1.0, 1.0)] #
              генерируем случайные числа с нормальным распределением
}
```

**4. Результаты.** Для сравнительной оценки точности полученного метода с базовым, таблично-гистограммным, выполнялось построение меры отклонения Кульбака-Лейблера при разных соотношениях  $k/N$  ( $k$  – размерность сетки интерполяции,  $N$  – количество отсчетов в синтетических данных, в данной работе принято равным 10 000) для каждого из 5 законов распределения. Результат сравнения показан на рис. 1. Законы распределения показаны в легенде к рис. 1: «**rrand**» – Релея, «**grand**» – Гамма, «**wrand**» – Вейбулла, «**exprand**» – экспоненциальное, «**gaussrand**» – нормальное (гауссово). Расчетные данные к рисунку 1 приведены в таблицах 1 и 2.



**Рис. 1.** Сравнительная точность интерполяции для таблично-гистограммного метода и для полученной нейросетевой интерполяции.

Таблица 1. Точность интерполяции гистограммным методом

Величина $k/N$	rrand	grand	wrand	exprand	gaussrand
0.0100	0.003333	-0.004176	0.004172	0.001169	0.001051
0.0166	0.005395	-0.002627	0.014493	0.002487	0.002485
0.0278	0.010528	-0.001127	0.027640	0.004309	0.005865
0.0464	0.019553	0.001291	0.065812	0.007118	0.009863
0.0774	0.033756	0.004193	0.106763	0.011950	0.018202
0.1291	0.063296	0.008189	0.186010	0.022943	0.029341
0.2154	0.120291	0.015558	0.398498	0.037571	0.051529
0.3593	0.226427	0.027862	0.706398	0.066140	0.107347
0.5994	0.483891	0.047202	1.495670	0.108961	0.198159
1.0000	1.012645	0.076647	3.077090	0.179611	0.404091

Таблица 2. Точность интерполяции при помощи нейросети

Величина $k/N$	rrand	grand	wrand	exprand	gaussrand
0.0100	0.216864	0.113760	0.442127	0.396204	0.065176
0.0166	0.318754	0.130621	0.576789	0.398202	0.103963
0.0278	0.327285	0.163094	0.657532	0.427052	0.131896
0.0464	0.557939	0.170919	0.826366	0.387221	0.142835
0.0774	0.389054	0.213921	0.884478	0.379168	0.157188
0.1291	0.493654	0.245398	0.910070	0.477201	0.174697
0.2154	0.468682	0.235149	0.931008	0.499094	0.178577
0.3593	0.454258	0.235967	0.965550	0.500662	0.188330
0.5994	0.531077	0.263226	0.985074	0.509339	0.179670
1.0000	0.413812	0.257140	0.997786	0.495032	0.185333

Как видно, в обоих случаях графики показывают монотонно возрастающие зависимости метрики отклонения от соотношения  $k/N$ , однако, нейросетевой интерполятор, обученный на данных численного дифференцирования функции распределения случайной величины, показывает на порядок меньшие значения метрики отклонения Кульбака-Лейблера между кривой оценки плотности распределения и теоретической кривой, чем базовый (гистограммный) метод, основанный на построении гистограммы.

**5. Вывод.** Предложенный подход к нейросетевой интерполяции функции плотности распределения случайной величины по заданному набору наблюдений отличается более высокой точностью, по сравнению с базовым таблично-гистограммным методом. Преимущество предложенного подхода особенно велико при больших значениях  $k/N$ , что позволяет говорить о преимуществах предложенного метода в ситуации малого объема экспериментальных данных, что особенно ценно для задачи анализа и прогнозирования бизнес-процессов, где частота выполнения операций редко превышает сотни, максимум – тысячи раз за период функционирования системы.

Таким образом, поставленная задача – построение нейросетевой интерполяции плотности распределения случайной величины – была успешно решена. Успех обеспечило совместно применение таких методов и алгоритмов, как:

1. Подготовка обучающих данных для нейросетевого интерполятора путем расчета значений функции плотности распределения случайной величины как первой производной от функции распределения.
2. Реализация интерполятора в виде многослойной нейронной сети (Sequential), позволяющей встраивать данную функцию в любой код, требующий получения значений функции плотности распределения случайной величины в заданных точках области распределения.

Дальнейшим развитием данной работы будет встраивание разработанной модели интерполятора в моделирующую часть цифрового двойника бизнес-процессов [17] и проверка его прогностической способности на реальных протоколах выполнения бизнес-процессов.

### Список источников

1. Ahmad T., Van Looy A. Business process management and digital innovations: a systematic literature review. Sustainability, 2020, vol. 12, no. 17, 6827 p.
2. Dorrer M. Implementation of the digital twin of educational business processes on the ELMA BPMS platform. Proceedings of the II International scientific conference on advances in science, engineering and digital education: (ASEDU-II 2021) AIP Conference Proceedings, 2022, 040088 p.
3. Hofacker I., Vetschera R. Algorithmical approaches to business process design. Comput. Oper. Res., 2001, vol. 28, no. 13, pp. 1253-1275.
4. Powell S.G., Schwaninger M., Trimble C. Measurement and control of business processes. Syst. Dyn. Rev., 2001, vol. 17, no. 1, pp. 63-91.
5. Valiris G., Glykas M. Business analysis metrics for business process redesign. Bus. Process Manag. J., 2004, vol. 10, no. 4, pp. 445-480.
6. Калянов Г.Н. Теория и практика реорганизации бизнес-процессов / Г.Н. Калянов. – Москва: СИНТЕГ, 2000. – 203 с.
7. Доррер М.Г. Оптимизация расчёта инвариантов сети Петри в рамках задачи формирования сценариев интеграционного тестирования / М.Г. Доррер, В.В. Курохтин // Моделирование и анализ информационных систем, 2012. – Т. 19. – № 4. – С. 128-143.
8. Петров М.И. Использование тензорной методологии анализа систем для описания бизнес-процессов предприятия. Средства автоматизации / М.И. Петров, И.Г. Треногий, Е.А. Беловатый // Сибирский аэрокосмический журнал, 2008. – Т. 3. – С. 157–161.
9. Kron G. Tensor analysis of networks. J. Wiley & Sons, 1939.
10. Barjis J., Dietz J.L.G. Business process modeling and analysis using GERT networks. Enterprise Information systems. Dordrecht: Springer Netherlands, 2000, pp. 71-80.
11. Zyryanov A.A., Dorrer M.G. The algorithm of business process model translation into the GERT-network model. Bull. KrasGAU, 2012, no. 12, pp. 13-18.
12. Ambika S. et al. Application of gert analysis in management. Int. J. Latest Eng. Manag. Res, 2018, vol. 03, no. 02, pp. 01-04.
13. Dorrer M., Dorrer A., Zyryanov A. Numerical modeling of business processes using the apparatus of GERT Networks. Society 5.0, Cyberspace for advanced human-centered society, 2021, pp. 47-55.
14. Phillips D.T., Garcia-Diaz A. Fundamentals of network analysis. Networks. Englewood Cliffs, NJ, Prentice Hall, 1981, 474 p.
15. Pritsker A.A.B. GERT: graphical evaluation and review technique. Santa Monica, CA: RAND Corporation, 1966. 150 p.
16. Доррер М.Г. Реализация цифрового двойника бизнес-процессов на базе системы ELMA / М.Г. Доррер // ИТНОУ информационные технологии в науке, образовании и управлении, 2021. – Т. 1 (17).
17. Dorrer M.G. The digital twin of the business process model. J. Phys. Conf. Ser., 2020, vol. 1679, 032096 p.
18. Dorrer M.G. The prototype of the organizational maturity model's digital twin of an educational institution. J. Phys. Conf. Ser., 2020, vol. 1691, 012121 p.
19. Шибанов А.П. Обобщенные GERT-сети для моделирования протоколов, алгоритмов и программ телекоммуникационных систем. / А.П.Шибанов // Рязанская государственная радиотехническая академия, 2003. – 307 с.
20. Зырянов А.А. Методика вероятностного прогнозирования состояния организационно-технологических систем при помощи формализмов GERT-сетей. / А.А. Зырянов // Сиб. аэрокосм. акад. им. акад. М.Ф. Решетнева, 2013. – 146 с.
21. Милащенко Е.А. Получаем кривую плотности распределения вероятности случайного процесса. Хабр, 2021. – URL: <https://habr.com/ru/articles/585232/>.
22. Прядченко В. Получаем кривую плотности распределения вероятности... быстрее и точнее. Хабр, 2021. – URL: <https://habr.com/ru/articles/587372/>.

*Доррер Михаил Георгиевич. К.т.н., СибГУ им. М.Ф. Решетнева, доцент. Направления исследований: искусственные нейронные сети, моделирование бизнес-процессов, имитационное моделирование. AuthorID: 114859, SPIN: 9983-0377, ORCID: 0000-0002-8598-8331, dorrer\_mg@sibsau.ru, 660037, Россия, г. Красноярск, пр. Мира, д.82*



UDC 004.852

DOI:10.25729/ESI.2024.34.2.006

## Evaluation of the quality of interpolation of the random variable distribution density using an artificial neural network

Mikhail G. Dorrer

Reshetnev Siberian State University of Science and Technology,

Russia, Krasnoyarsk, *dorrer\_mg@sibsau.ru*

**Abstract.** The paper describes the creation and evaluation of the performance of a neural network model of the probability distribution density function of a random variable, given by a set of measurements of a random variable in the absence of the identification stage of the distribution law. The need to solve this problem is caused by limitations introduced into the accuracy of calculating the probability distribution density function of a random variable both by the tabular-histogram method and in the case of applying approaches to the identification of the distribution law. The problem was solved in Python using the TensorFlow neural network library by creating a neural network model based on the Sequential class with fully connected Dense layers, trained on data from numerical differentiation of the random variable distribution function. The accuracy of the forecast was estimated using the Kullback-Leibler distance measure for various ratios of the volume of experimental data and the number of interpolation intervals on synthetic test data generated for 5 laws of distribution - Rayleigh, Weibull, gamma, exponential and normal (Gaussian). To assess the predictive ability of the approach when testing the interpolator, random variable samples shifted relative to those used in training were used. The proposed solution showed a significantly higher accuracy in calculating the values of the distribution density of a random variable compared to the histogram method. The developed approach will be implemented in the modeling part of the digital twin of a business process based on the mathematical apparatus of stochastic GERT networks.

**Keywords:** stochastic models, random variable distribution density, neural networks, digital process models

### References

1. Ahmad T., Van Looy A. Business process management and digital innovations: a systematic literature review. *Sustainability*, 2020, vol. 12, no. 17, 6827 p.
2. Dorrer M. Implementation of the digital twin of educational business processes on the ELMA BPMS platform. *Proceedings of the II International scientific conference on advances in science, engineering and digital education: (ASEDU-II 2021) AIP Conference Proceedings*, 2022, 040088 p.
3. Hofacker I., Vetschera R. Algorithmical approaches to business process design. *Comput. Oper. Res.*, 2001, vol. 28, no. 13, pp. 1253-1275.
4. Powell S.G., Schwaninger M., Trimble C. Measurement and control of business processes. *Syst. Dyn. Rev.*, 2001, vol. 17, no. 1, pp. 63-91.
5. Valiris G., Glykas M. Business analysis metrics for business process redesign. *Bus. Process Manag. J.*, 2004, vol. 10, no. 4, pp. 445-480.
6. Kalyanov G.N. *Teoriya i praktika reorganizatsii biznes-protsessov [Theory and practice of business process reorganization]*. Moscow: SINTEG, 2000. 203 p.
7. Dorrer M.G., Kurohtin V.V. Optimizatsiya rascheta invariantov seti Petri v ramkakh zadachi formirovaniya skenarijev integratsionnogo testirovaniya [Optimization of the calculation of Petri net invariants within the task of forming scenarios of integration testing]. *Modelirovanie i analiz informatsionnykh sistem [Modeling and analysis of information systems]*, 2012, vol. 19, no. 4, pp. 128-143.
8. Petrov M.I., Trenogiy I.G., Belovaty E.A. Ispol'zovanie tenzornoy metodologii analiza sistem dlya opisaniya biznes-protsessov predpriyatiya. Sredstva avtomatizatsii [Using tensor methodology for system analysis to describe enterprise business processes. Automation tools]. *Sibirskiy aerokosmicheskiy zhurnal [Siberian aerospace journal]*, 2008, vol. 3, pp. 157-161.
9. Kron G. *Tensor analysis of networks*. J. Wiley & Sons, 1939.
10. Barjis J., Dietz J.L.G. *Business process modeling and analysis using GERT networks*. Enterprise Information systems. Dordrecht: Springer Netherlands, 2000, pp. 71-80.
11. Zyryanov A.A., Dorrer M.G. The algorithm of business process model translation into the GERT-network model. *Bull. KrasGAU*, 2012, no. 12, pp. 13-18.
12. Ambika S. et al. Application of gert analysis in management. *Int. J. Latest Eng. Manag. Res*, 2018, vol. 03, no. 02, pp. 01-04.

13. Dorrer M., Dorrer A., Zyryanov A. Numerical modeling of business processes using the apparatus of GERT Networks. Society 5.0, Cyberspace for advanced human-centered society, 2021, pp. 47-55.
14. Phillips D.T., Garcia-Diaz A. Fundamentals of network analysis. Networks. Englewood Cliffs, NJ: Prentice Hall, 1981, 474 p.
15. Pritsker A.A.B. GERT: graphical evaluation and review technique. Santa Monica, CA: RAND Corporation, 1966. 150 p.
16. Dorrer M.G. Realizatsiya tsifrovogo dvoynika biznes-protsessov na baze sistemy ELMA [Implementation of the digital twin of business processes on the ELMA system]. ITNOU informatsionnye tekhnologii v nauke, obrazovanii i upravlenii [ITNOU information technologies in science, education and management], 2021, vol. 1 (17).
17. Dorrer M.G. The digital twin of the business process model. J. Phys. Conf. Ser., 2020, vol. 1679, 032096 p.
18. Dorrer M.G. The prototype of the organizational maturity model's digital twin of an educational institution. J. Phys. Conf. Ser., 2020, vol. 1691, 012121 p.
19. Shibanov A.P. Obobshchennye GERT-seti dlya modelirovaniya protokolov, algoritmov i programm telekommunikatsionnykh sistem [Generalized GERT networks for modeling protocols, algorithms and programs of telecommunication systems]. Ryazanskaya gosudarstvennaya radiotekhnicheskaya akademiya [Ryazan State radio engineering academy], 2003, 307 p.
20. Zyryanov A.A. Metodika veroyatnostnogo prognozirovaniya sostoyaniya organizatsionno-tekhnologicheskikh sistem pri pomoshchi formalizmov GERT-setey [Methodology for probabilistic forecasting of the state of organizational and technological systems using GERT network formalisms]. Sibirskiy aerokosmicheskii akademiya im. akad. M.F. Reshetneva [Siberian aerospace academy named after M.F. Reshetnev], 2013, 146 p.
21. Milashchenko E.A. Poluchaem krivuyu plotnosti raspredeleniya veroyatnosti sluchaynogo protsessa [Getting the probability density curve of a random process]. Habr, 2021, available at: <https://habr.com/ru/articles/585232/>.
22. Pryadchenko V. Poluchaem krivuyu plotnosti raspredeleniya veroyatnosti... bystree i tochnee [Getting the probability density curve... faster and more accurate]. Habr, 2021, available at: <https://habr.com/ru/articles/587372/>.

**Dorrer Mikhail Georgievich.** Ph.D., Reshetnev university, docent. Research direction: artificial neural networks, business process modeling, simulation modeling. AuthorID: 114859, SPIN: 9983-0377, ORCID: 0000-0002-8598-8331, dorrer\_mg@sibsau.ru, 660037, Russian Federation, Krasnoyarsk, Krasnoyarsky Rabochy Ave.

*Статья поступила в редакцию 08.12.2023; одобрена после рецензирования 30.05.2024; принята к публикации 03.06.2024.*

*The article was submitted 12/08/2023; approved after reviewing 05/30/2024; accepted for publication 06/03/2024.*