

УДК 004.032.26:519.248

## ИСКУССТВЕННАЯ НЕЙРОННАЯ СЕТЬ ДЛЯ ПРОГНОЗИРОВАНИЯ ЧИСЛЕННОСТИ НАСЕЛЕНИЯ

Носова М.Г., Фёдоров А.В.

ФГБОУ ВО «Томский государственный университет систем управления и радиоэлектроники»,  
Томск, e-mail: nosovamgm@gmail.com

В данной статье представлена разработка и исследование модели для прогнозирования численности населения в виде искусственной нейронной сети. В процессе исследования проводится выбор оптимальных средств и инструментов создания и моделирования нейронной сети, выполняется обучение и оптимизация модели нейронной сети. Для искусственной нейронной сети произведена оптимизация основных технологических характеристик: точность, соотношение тренировочного и обучающего множеств, количество эпох, архитектура сети. При решении данной задачи использованы данные о численности населения Российской Федерации за период с 1960 по 2019 г., на основании которых сделан прогноз до 2036 г. Наиболее высокую точность при высокой скорости обучения нейронной сети на тестовых данных показывает следующая модель искусственной нейронной сети: функция активации ReLU, количество скрытых слоев нейронной сети 5 и количество нейронов 100. Сеть обучилась достаточно успешно, поэтому полученные с ее помощью данные о численности населения Российской Федерации на период с 2019 по 2036 г. можно считать в той или иной степени корректными. Из чего следует, что разработанная нейронная искусственная сеть вполне пригодна для прогнозирования численности населения на краткосрочный период.

**Ключевые слова:** искусственный интеллект, нейронная сеть, численность населения, прогнозирование, демография

## ARTIFICIAL NEURAL NETWORK FOR POPULATION FORECASTING

Nosova M.G., Fyodorov A.V.

Tomsk State University of Control Systems and Radioelectronics, Tomsk, e-mail: nosovamgm@gmail.com

This article presents the development and study of a model for predicting population size in the form of an artificial neural network. In the research, the choice of optimal means and tools for creating and modeling a neural network is carried out, training and optimization of the neural network model is performed. For an artificial neural network, the optimization of the main technological characteristics has been made: accuracy, the ratio of training and training sets, the number of epochs, and the architecture of the network. This task uses data of the population of the Russian Federation for the period from 1960 to 2019, on the basis of which a forecast is made until 2036. The fastest neural network learning rate on test data is shown by the following model of an artificial neural network: the activation function ReLU, the number of hidden layers of the neural network is 5 and the number of neurons is 100. The network has learned quite successfully, so that the data obtained with its help on the population of the Russian Federation for the period from 2019 to 2036 can be considered correct to one degree or another. It follows from this that the developed artificial neural network is quite suitable for predicting the population size in the short term.

**Keywords:** artificial intelligence, neural network, population, forecasting, demography

В настоящее время востребованными являются перспективные расчеты общей численности населения. Результаты демографического прогноза необходимы при планировании в экономических и социальных областях, например в образовании, производстве, здравоохранении, спорте и т.д. В связи с этим задача построения научно обоснованных демографических прогнозов, как на ближайшее, так и на более отдаленное будущее, является особенно актуальной [1].

Для прогнозирования численности населения Российской Федерации используются различные статистические методы [2–4] и методы математического моделирования [5–7]. Поскольку численность населения зависит от влияния ряда факторов, ее прогнозирование является безусловной сложной задачей. Изменение в одном факторе может сильнее или слабее повлиять на численность населения в различное вре-

мя. Поэтому численность населения можно считать хаотичной системой [8].

Официальные демографические прогнозы в нашей стране на регулярной основе публикуются с 1993 г. [9]. Однако их реалистичность и степень определенности вызывают дискуссии в научных кругах, поскольку используемый, как правило, статистический подход плохо подходит для прогнозирования хаотических систем, таких как численность населения.

Многообразие статистических и математических моделей доказывает сложность построения общей модели процесса изменения демографической ситуации, которая включала бы в себя влияние всех общественных, экономических и биологических факторов [2, 4, 6]. Кроме того, в современных условиях необходимость развития методов и моделей прогнозирования усиливается неоднозначностью демографических трендов последних лет даже в сравнительно недолго-

срочной перспективе. В связи с этим проблема разработки новых математических моделей процесса изменения демографической ситуации, а также методов их исследования, является весьма важной и актуальной.

Одно из последних направлений в науке представлено искусственными нейронными сетями [10]. Известно об успешном применении искусственных нейронных сетей для прогнозирования в различных направлениях статистики [10]. Искусственные нейронные сети (ИНС) полезны в моделях, где базовые процессы имеют хаотичный характер. Поэтому, учитывая нелинейность данных численности населения, существует интерес в применении искусственных нейронных сетей для прогнозирования численности населения.

Целью данной работы является разработка новой математической модели для прогнозирования численности населения в виде искусственной нейронной сети, а также применение полученных результатов к исследованию процесса изменения численности населения в Российской Федерации.

*Основные параметры модели нейронной сети для прогнозирования численности населения*

ИНС представляет собой математическую модель и способна воспринимать информацию как биологический нейрон [11]. Основные компоненты нейронных сетей можно разделить на архитектуру и функциональные свойства. Архитектура определяет структуру сети, то есть количество искусственных нейронов в сети и их взаимосвязь. Функциональные свойства ИНС определяют способ обучения и получения прогноза. Фундаментальными исследованиями в области нейронных сетей занимался С. Хайкин. В его трудах приводятся математическое обоснование нейросетевых алгоритмов, примеров и описание компьютерных экспериментов с нейронными сетями [10].

В рамках проектирования искусственной нейронной сети был выбран язык программирования Python и фреймворк Keras. Keras обладает теми утилитами, которые позволяют автоматизировать этот процесс.

Для подбора оптимальной модели искусственной нейронной сети будут протестированы следующие параметры:

– протестированы следующие функции активации: линейная, сигмоидальная, гиперболический тангенс и Rectified Linear Unit (ReLU);

– изменение глубины сети (количества скрытых слоев) с фиксированным количеством нейронов в каждом слое;

– изменение ширины сети (количества нейронов) с фиксированным количеством слоев.

Эксперименты с изменением глубины нейросети будут проведены по следующим параметрам:

– количество скрытых слоев изменялось с 2 до 6;

– количество нейронов в скрытых слоях фиксировано 50.

Эксперименты с изменением ширины нейросети проводились по следующим параметрам:

– количество нейронов в скрытых слоях изменялось на 10, 50, 100, 200, 500;

– количество скрытых слоев фиксировано 4.

Для определения численности населения необходимы данные о численности населения за прошлые годы с сайта World Bank Data [12]. В данных представлена статистика о численности населения с 1960 по 2019 г. по странам. В качестве тестирования используются данные о численности населения Российской Федерации (РФ).

В качестве проверки точности прогноза будет проведено сравнение полученных результатов с официальной статистикой численности населения от Федеральной службы государственной статистики (Росстат) с 2011 по 2019 г. по методу скорректированного квадратичного отклонения (adjusted R Squared) [9, 11].

*Реализация модели искусственной нейронной сети*

Для реализации ИНС для прогнозирования численности населения произведен анализ доступных библиотек машинного обучения. Среди рассматриваемых библиотек выделены Tensorflow, Keras, PyTorch, Caffe, Theano, Caffe, DeepLearning4J. Для выбора эффективной библиотеки рассмотрена задача классификации изображений, как самый распространенный тест эффективности библиотек машинного обучения. В анализе для выбранных библиотек реализована рекуррентная модель LSTM-типа ИНС с различными параметрами для оценки эффективности построения модели с помощью различных программных средств. Как уже говорилось ранее, в качестве оптимальной библиотеки выбрана Keras. Keras выгодно выделяется упрощением многих аспектов разработки ИНС, широкими настройками ИНС, совмещение в себе нескольких библиотек машинного обучения. В качестве языка программирования выбран Python. При реализации ИНС также использованы библиотеки Pandas, matplotlib, NumPy, Scikit-Learn и облачный сервис Google Colaboratory.

Для начала работы ИНС необходимо получить последние данные о численности населения с World Bank Data [12], предоставляющего csv файл с актуальными данными. С помощью методов `drop()`, `transpose()` и `index()` реализован поиск данных по стране Российская Федерация и установка индексов элементов на год. Федеральная служба государственной статистики предоставляет прогноз до 2035 г. в файле формата excel, поэтому получение данных из файла реализуется функцией `read_excel()`. Пример реализации указанных методов при вводе данных показан на рис. 1.

Делим полученные данные на тестовые и обучаемые, определив для них 2 массива и масштабировав от 0 до 1 для корректной работы вычислительных алгоритмов. Для этого используются методы `MinMaxScaler fit_transform()` из библиотеки `Scikit-learn`. Пример масштабирования данных с помощью метода `MinMaxScaler` показан на рис. 2.

Следующим шагом составим модель ИНС с помощью библиотеки `Keras` и следующих методов:

- `Sequential()` – обозначает последовательную нейронную сеть;
- `Dense()` объекта `layers`, определяет параметры добавляемого скрытого слоя, среди параметров выделяются количество слоев, функция активации;
- `Compile()` – представляет собой настройку процесса обучения ИНС, содер-

жит определение функции ошибки, которая минимизируется в процессе обучения (например, среднеквадратичная ошибка), и оптимизатора, который определяет процедуру обучения, можно выделить оптимизатор `Adam`;

– `Fit()` – представляет собой реализацию обучения ИНС, определяет количество итераций обучения по всем входным данным (эпохи) и разбиение данных на меньшие блоки (пакеты), обрабатываемых моделью одновременно, в процессе обучения один пакет используется в градиентном спуске для вычисления изменения весов модели;

– `Predict()` – метод используется для получения прогнозируемых данных из ИНС после обучения сети.

Пример реализации модели ИНС с помощью методов библиотеки `Keras` показан на рис. 3.

Для оценки точности прогнозирования численности населения созданных моделей нейронной сети используется архитектура рекуррентной нейронной сети LSTM-типа. Для оценки масштабируемости рассмотренных библиотек расчет показателей производился для различных коэффициентов масштабирования сетей. Обучение искусственной нейронной сети производится на 80% тестовых данных, а для проверки точности прогноза – оставшиеся 20% данных. В качестве показателя точности используется коэффициент детерминации.

```
df = pd.read_csv('/content/country_population.csv')
ds = df[df['Country Name'] == 'Russian Federation']
ds = ds.reset_index()
ds = ds.drop(['Country Code', 'index', 'Indicator Code', 'Indicator Name', 'Country Name'], axis=1)
```

Рис. 1. Скриншот получения данных о численности населения в РФ

```
from sklearn.preprocessing import MinMaxScaler
sc = MinMaxScaler()
train_sc = sc.fit_transform(train1)
test_sc = sc.transform(test1)
```

Рис. 2. Масштабирование данных методом `MinMaxScaler`

```
K.clear_session()
model1 = Sequential()
model1.add(Dense(50, input_shape=(X_test1ds.shape[1],), activation='relu', kernel_initializer='lecun_uniform'))
model1.add(Dense(50, input_shape=(X_test1ds.shape[1],), activation='relu', kernel_initializer='lecun_uniform'))
model1.add(Dense(50, input_shape=(X_test1ds.shape[1],), activation='relu', kernel_initializer='lecun_uniform'))
model1.add(Dense(1))
model1.compile(optimizer=Adam(lr=0.001), loss='mean_squared_error')
model1.fit(X_train1ds, y_train1ds, batch_size=40, epochs=200, verbose=1)
```

Рис. 3. Скриншот модели ИНС

При исследовании оптимальных параметров модели нейронной сети для прогнозирования численности населения наиболее точный прогноз на тестовых данных показала следующая модель искусственной нейронной сети: функция активации ReLU, количество скрытых слоев нейронной сети 5 и количество нейронов 100.

*Сравнение точности прогноза реализованной искусственной нейронной сети для прогнозирования численности населения Российской Федерации*

Для того чтобы сказать, что построенный с помощью нейронной искусственной сети прогноз численности населения правилен и адекватен, необходимо, чтобы рассчитанные нами значения численности населения с 2011 по 2018 г. были наиболее близки к статистическим данным Федеральной службы государственной статистики по этим годам [9]. Росстат осуществляет функцию по формированию официальной статистической информации о социальных, экономических, демографических и других процессах в РФ, а также в порядке и случаях, установленных законодательством РФ по контролю в сфере официального статистического учета. Методология прогнозирования демографической ситуации в РФ основывается на множестве показателей. Среди показателей, влияющих на официальную статистику, выделяются экономический достаток населения, областей проживания, множество показателей смертности населения от различных факторов (болез-

ней, дорожно-транспортных происшествий и других), рождаемость, доля трудоспособного населения. Такое множество показателей требует больших затрат на составление формулы расчета численности населения, но обеспечивает основательную статистику. Сравнение результатов прогнозирования искусственной нейронной сети с реальными данными по численности населения позволит оценить перспективность расширения ИНС с включением большего числа показателей для увеличения точности расчета численности населения в РФ.

Представим на графике (рис. 4) данные по реальной численности населения согласно официальной статистике с 2011 по 2018 г. [9], данные прогноза Росстата за 2010 г. на этот период [9] и результаты прогноза численности, полученные с помощью ИНС. Нейронная сеть обучалась по данным до 2010 г.

Как видно из графика, прогноз ИНС до 2014 г. близок к реальным данным по численности населения (коэффициент детерминации 0,968). Однако в 2014 г. в состав Российской Федерации включен Крым, этим объясняется неожиданный рост численности населения на графике. Из-за включения в состав РФ Крыма следующая часть графика имеет низкую точность. Для решения подобных проблем необходимо сокращать дальность прогноза либо исключать подобные факторы.

На рис. 5 представлены результаты прогноза численности, полученные с помощью ИНС, и результаты прогноза численности населения Росстата до 2035 г. [9].

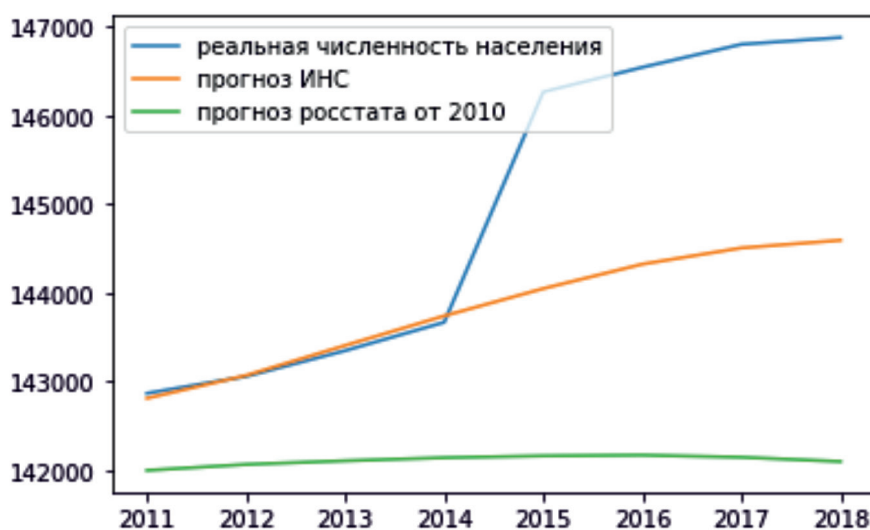


Рис. 4. График сравнения прогноза ИНС, официальной статистики численности населения РФ и прогноза Росстата

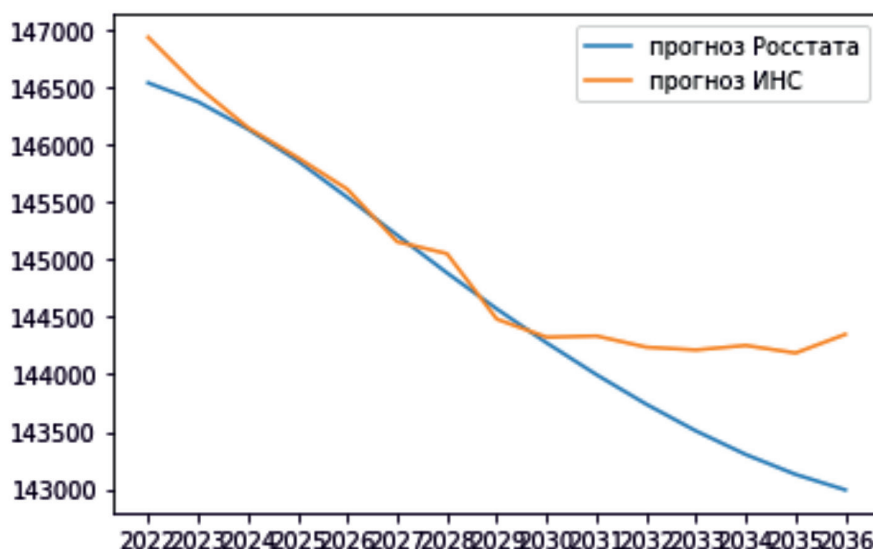


Рис. 5. График сравнения прогноза нейросети

Расхождение данных связано с пандемией COVID-19, это вносит свои изменения в график прогноза. В первые годы график прогноза ИИС соответствует графику прогноза Росстата, но после 2030 г. наблюдается расхождение.

### Заключение

В результате выполнения данной работы была спроектирована и обучена нейронная сеть, которая была применена для прогнозирования численности населения Российской Федерации. Сеть обучилась достаточно успешно, поэтому полученные с ее помощью данные о численности населения Российской Федерации на период с 2019 по 2036 г. можно считать в той или иной степени корректными. Из чего следует, что разработанная нейронная искусственная сеть вполне пригодна для прогнозирования численности населения на краткосрочный период.

Опираясь на полученные в данной работе результаты, в дальнейшем планируется расширение ИИС с включением большего числа показателей для увеличения точности прогноза численности населения в РФ.

### Список литературы

1. Косов П.И., Берендеева А.Б. Основы демографии. М.: Инфра-М, 2014. 288 с.
2. Носова М.Г. Автономная немарковская система массового обслуживания и ее применение в задачах демографии: дис. ... канд. физ.-мат. наук. Томск, 2010. 204 с.
3. Girosi F., King G. Demographic Forecasting, Princeton: Princeton University Press, 2018. DOI: 10.1515/9780691186788.
4. Yarnykh E.A., Oleynikov B.I., Tsylyna I.T. Statistical methods for the analysis of demographic processes within the region. 8th International Scientific Conference Science and Society. 2015. P. 117–123.
5. Носова М.Г. Стохастическая модель для прогнозирования численности населения // Бюллетень науки и практики. 2019. Т. 5. № 9. С. 17–25.
6. Nosova M. A Mathematical Model of Population Growth as a Queuing System. arXiv preprint arXiv:2005.10518. 2020 May 21.
7. Hinde A. Demographic methods. Routledge, 2014. 320 p.
8. Антонова Н.Л. Демография. Екатеринбург: Изд-во Урал. ун-та, 2014. 154 с.
9. Демография // Федеральная служба государственной статистики [Электронный ресурс]. URL: <https://rosstat.gov.ru/folder/12781> (дата обращения: 03.06.2021).
10. Хайкин С. Нейронные сети. М.: ООО «И.Д. Вильямс», 2016. 1104 с.
11. Ростовцев В.С. Искусственные нейронные сети: учебник. М.: Изд-во «Лань», 2019. 216 с.
12. Population // Data – World Bank Data – World Bank Group [Electronic resource]. URL: <https://data.worldbank.org/indicator/SP.POP.TOTL> (date of access: 03.06.2021).