

О. В. Лимановская

*Уральский федеральный университет
им. первого Президента России Б. Н. Ельцина*

Екатеринбург, Россия

И. В. Гаврилов

В. Н. Мещанинов

Уральский государственный медицинский университет

Екатеринбург, Россия

Применение нейронных сетей для повышения объема обучающей выборки

Предложен механизм получения искусственной выборки медицинских численных данных методами искусственного интеллекта. Разработанная генеративно-сопоставительная нейронная сеть позволяет получить выборку медицинских данных в любом заданном пользователем системы объеме на основе имеющихся функциональных данных пациента. В работе получена искусственная выборка функциональных данных пациента объемом 100 тыс. записей по 11 параметрам (возраст, АДС, АДД, АДП, задержка дыхания на вдохе, задержка дыхания на выдохе, жизненная емкость легких, масса тела, аккомодация, острота слуха, статическая балансировка). Полученная выборка использовалась для построения прогностической модели определения биовозраста с помощью нейронных сетей. Средняя абсолютная ошибка полученной модели составила три года, в то время как средняя абсолютная ошибка моделей, полученных на исходных выборках, составила пять лет.

Ключевые слова: нейронные сети, биовозраст, генерация данных, прогнозирование биовозраста

Oksana V. Limanovskaya

*Ural Federal University named
after the first President of Russia B. N. Yeltsin*

Yekaterinburg, Russia

Iliya V. Gavrilov

Viktor N. Meshchaninov

Ural State Medical University

Yekaterinburg, Russia

The Use of Neural Networks to Increase the Volume of the Training Sample

A mechanism for obtaining artificial sampling of medical numerical data by artificial intelligence methods is proposed. The developed generative-adversarial neural network allows you to get a sample of medical data in any volume specified by the user of the system based on the available functional data of the patient. In the work, an artificial sample of the patient's functional data with a volume of 100,000 records was obtained for 11 parameters (age, ADS, ADD, ADP, breath retention on inhalation, breath retention on exhalation, vital lung capacity, body weight, accommodation, hearing acuity, static balancing). The resulting sample was used to build a predictive model for determining bio-age using neural networks. The average absolute error of the obtained model was 3 years, while the average absolute error of the models obtained from the initial samples was 5 years.

Keywords: neural networks, bio-growth, data generation, prediction of bio-growth

Введение. Дефицит медицинских данных при построении прогнозных моделей ограничивает круг возможных технологий построения математических моделей и уменьшает точность получаемых моделей. Возможность увеличения объема выборки данных методами искусственного интеллекта позволит на основе имеющихся небольших выборок медицинских данных получать выборки, пригодные для построения более точных моделей с использованием нейронных сетей. В работах [1; 2] показана возможность использования генеративно-сопоставительных нейронных сетей для создания

искусственных наборов данных и увеличения имеющихся объемов выборки. Механизм генеративно-состязательных нейронных сетей позволяет получать данные, сохраняющие начальное статистическое распределение данных, что дает возможность использовать полученные выборки для обучения моделей машинного обучения и нейронных сетей. Настоящая работа посвящена разработке генеративно-состязательной нейронной сети для получения выборки данных на основе имеющегося набора функциональных данных пациентов.

Материалы и методы. Основой для генерации данных послужила выборка, полученная в медицинской организации ГАУЗ СО «СОКП Госпиталь для ветеранов войн» за 1995–2014 гг. в объеме 536 записей, включающая в себя функциональные показатели пациентов (11 параметров), впервые пришедших на обследование. Построение генеративно-состязательной нейронной сети проводилось на платформе *Anaconda* с дистрибутивом *Python 3.6* [3] с использованием библиотеки *tensorflow* [4].

Результаты. Разработана полносвязная генеративно-состязательная нейронная сеть, состоящая из генератора и дискриминатора. Генератор создает данные с распределением, максимально близким к истинному распределению реальных данных на основе поданных в него случайных данных. Полученные им данные отправляются на вход дискриминатору, который классифицирует их на те, что получены из реального набора данных, и на искусственные. Генератор и дискриминатор представляют собой отдельные нейронные сети, и их совместная работа продолжается до тех пор, пока генератор не научится создавать данные, настолько похожие на истинные, что дискриминатор не будет способен их отличить. После этого используется только генератор.

В разработанной модели генеративно-состязательной нейронной сети генератор представлен трехслойной полносвязной нейронной сетью. В первом слое использовано 512 нейронов, во втором — 252 нейрона, в третьем — 20 нейронов.

Дискриминатор представлен трехслойной полносвязной нейронной сетью. В первом слое использовано 512 нейронов, во втором — 252 нейрона, в третьем — 10 нейронов.

Обучение генеративно-сопоставительной нейронной сети велось на наборе функциональных данных пациентов объемом 536 записей из 11 параметров (возраст, АДС, АДД, АДП, задержка дыхания на вдохе, задержка дыхания на выдохе, жизненная емкость легких, масса тела, аккомодация, острота слуха, статическая балансировка). В результате получена выборка данных объемом 100 тыс. записей для этих же параметров.

Полученная выборка использовалась для обучения полносвязной нейронной сети, построенной для прогнозирования биовозраста на основе функциональных данных пациента. Средняя абсолютная ошибка полученной модели составила 3,7 года, что улучшило прежний результат прогнозирования, полученный в работе [5].

Заключение. В результате исследования получена генеративно-сопоставительная нейронная сеть, способная создавать наборы искусственных данных. Проведено ее испытание, и модель, обученная на полученном наборе данных, показала более высокую точность, чем построенная ранее модель на исходных данных.

Библиографические ссылки

1. *Liu L., Xia Yu. et al.* An overview of biological data generation using generative adversarial networks // 2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS) : [electronic resource]. 6–9 July. 2020. P. 141–144. URL: <https://www.cosmoc.org> (accessed: 28.10.2021).

2. *Alqahtani H., Kavakli-Thorne M. et al.* Applications of Generative Adversarial Networks (GANs): An Updated Review // Archives of Computational Methods in Engineering. 2019. Springer. P. 525–552.

3. *Anaconda* : [website]. URL: <https://www.anaconda.com> (accessed: 18.10.2021).

4. *TensorFlow* : [website]. URL: <https://www.tensorflow.org> (accessed: 28.10.2021).

5. *Лимановская О. В., Гаврилов И. В. и др.* Моделирование биологического возраста пациентов на основе их функциональных показателей // Моделирование, оптимизация и информационные технологии : [сайт]. URL: <https://moitvvt.ru/ru/journal/pdf?id=966> (дата обращения: 04.11.2021).