

УДК 004.934.2

Научная статья

<https://doi.org/10.32603/2071-8985-2022-15-5/6-71-79>

Распознавание речевого сигнала монгольского языка по слогам

Б. Зандан^{1✉}, Т. Галбаатар¹, О. Бухцоож^{2,3}, А. Г. Ченский⁴

¹ Институт физики и технологии Академии наук Монголии, Улан-Батор, Монголия

² Институт математики и цифровых технологий Академии наук Монголии, Улан-Батор, Монголия

³ Национальный исследовательский университет «МЭИ», Москва, Россия

⁴ Иркутский национальный исследовательский технический университет, Иркутск, Россия

zandan1984@gmail.com✉

Аннотация. Рассмотрена возможность распознавания речи монгольского языка с использованием нейронной сети. Применен способ с выделением слога распознаваемого слова, который позволяет уменьшить размер базы данных, поскольку при этом нет нужды сохранять признаки каждого слова. Обучение нейронной сети выполнено алгоритмом Левенберга–Марквардта, более устойчивым и быстродействующим при работе с малоразмерной базой данных. В эксперименте использовали 4 слова монгольского языка с одним или двумя слогами, при записи использовалось программное обеспечение MatLab. Слова озвучены 11 дикторами, мужчинами и женщинами разного возраста: 1 человек младше 30 лет, 7 человек в возрасте 30–40 лет и 3 человека – 60–70 лет. Кроме того, нейронная сеть распознает участок в речи, где идет шум, чтобы различать и отделить речевой участок от шума. Обученная сеть распознает слова монгольского языка по слогам из базы данных с вероятностью 96.5 %.

Ключевые слова: автоматическое распознавание речи, алгоритм Левенберга–Марквардта, коэффициент MFCC, искусственная нейронная сеть, слог слова

Для цитирования: Распознавание речевого сигнала монгольского языка по слогам / Б. Зандан, Т. Галбаатар, О. Бухцоож, А. Г. Ченский // Изв. СПбГЭТУ «ЛЭТИ». 2022. Т. 15, № 5/6. С. 71–79. doi: 10.32603/2071-8985-2022-15-5/6-71-79.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Original article

Recognition of the speech signal of the Mongolian language by syllables

B. Zandan^{1✉}, T. Galbaatar¹, O. Bukhtsooj^{2,3}, A. G. Chensky⁴

¹ Institute of Physics and Technology of the Academy of Sciences of Mongolia, Ulan Bator, Mongolia

² Institute Mathematics and Digital technologies of the Academy of Sciences of Mongolia, Ulan Bator, Mongolia

³ MPEI, Moscow, Russia

⁴ Irkutsk National Research Technical University, Irkutsk, Russia

zandan1984@gmail.com✉

Abstract. In this paper, we considered the possibility of Mongolian speech recognition using a neural network. At the same time, we applied a method with highlighting the syllable of a recognizable word, since in the Mongolian language words are formed by adding a suffix to its root. And the method of extracting the syllable of a word allows you to reduce the size of the database, since there is no need to save the features of each word.

The neural network was trained using the Levenberg–Marquardt algorithm, which is more stable and faster when working with a small database. The experiment used 4 words of the Mongolian language with one or two syllables. The words were written directly into the computer using the Matlab software. The speakers who voiced these words are 11 people, men and women of different ages. Of these, 1 person aged 20–30 years, 7 people aged 30–40 years and 3 people aged 60–70 years. In addition, the neural network recognizes the area in speech where the noise occurs in order to distinguish and separate the speech area from the noise. The trained network recognizes the words of the Mongolian language by syllables from the database with a probability of 96.5 %.

Keywords: automatic speech recognition, Levenberg–Marquardt algorithm, MFCC coefficient, artificial neural network, word syllable

For citation: Zandan B., Galbaatar T., Bukhtsooj O., Chensky A. G. Recognition of the speech signal of the Mongolian language by syllables // LETI Transactions on Electrical Engineering & Computer Science. 2022. Vol. 15, no. 5/6. P. 71–79. doi: 10.32603/2071-8985-2022-15-5/6-71-79.

Conflict of interest. The authors declare no conflicts of interest.

Введение. Автоматическое распознавание речи остается активной областью изучения взаимодействия человека с машиной. В последние годы речевые технологии изменили наш образ жизни и стали основным способом взаимодействия людей с устройствами. В настоящее время многие технологии распознавания речи успешно разработаны и используются на практике (Apple Siri, Amazon Alexa, Google Cloud Speech, Яндекс Алиса). В последнее время возрастает количество исследований по распознаванию монгольской речи, и первое интернет-приложение для распознавания речи (Chimege), использующее облачные технологии, было представлено общественности. Однако существует необходимость продолжить изучение распознавания монгольского языка, чтобы улучшить способность распознавать длинные слова, улучшить распознавание в шумной среде и разработать независимое устройство распознавания речи.

Речь распознается по словам, слогам и звукам. В [1] мы успешно провели эксперимент по распознаванию речи монгольского языка. Хотя распознавание речи с помощью слов эффективно, но в монгольском языке слова образуются добавлением суффиксов к корню слова [2]. Таким образом, у слов один и тот же корень, но разные суффиксы, и каждое слово необходимо хранить в отдельной библиотеке. В этом исследовании был проведен эксперимент по распознаванию речи с выделением слогов в словах с использованием искусственной нейронной сети.

Процесс распознавания речи. Процесс распознавания речи выполняется в порядке предварительной обработки речевых сигналов, выделе-

ния признаков и классификации [3]. Предварительная обработка речевого сигнала выполняется, как показано на рис. 1.



Рис. 1. Последовательность предварительной обработки речевого сигнала

Fig. 1. Speech signal preprocessing sequence

Предварительная обработка речевого сигнала. Речевой сигнал, записанный через микрофон, преобразуется в цифровой сигнал с разрешением 16/44.1 бит/кГц с помощью звуковой карты компьютера и программного пакета MatLab. Далее речевой сигнал был нормирован для выравнивания амплитуды сигнала каждой записи. Балансировка отношения низких и высоких частот в сигнале улучшает распознавание речевого сигнала, и для этого применялся КИХ-фильтр. Речевой сигнал обычно нестабилен, однако его можно считать относительно стабильным на короткий промежуток времени ~10...30 мс [4]. Следовательно, сигнал обрабатывается разбиением его на короткие кадры с использованием оконной функции. Первые пять шагов предварительной обработки сигнала выполняются согласно [5]. На рис. 2 показан результат предварительной обработки сигнала после 5-го этапа. Амплитуда сигнала на этом рисунке нормирована согласно схеме на рис. 1, но не равна единице, поскольку сигнал проходит через фильтр. Далее на других ри-

сунках эпюры – это также нормированные амплитуды, не равные единице, т. е. сигналы проходят через фильтр на разных этапах их обработки. Аналоговый речевой сигнал для обработки должен быть дискретизован. Сигнал представляют совокупностью его значений, называемых выборками, взятыми в период времени, обычно равный периоду дискретизации.

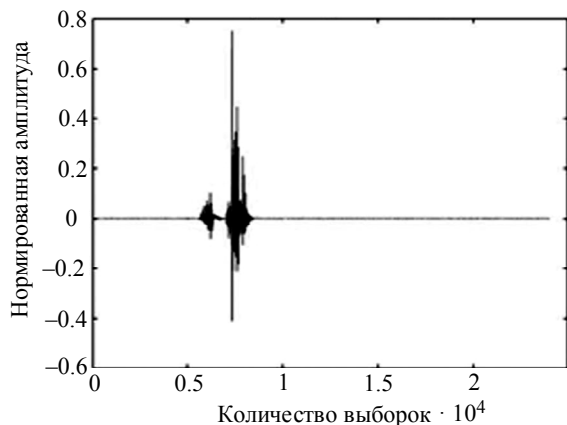


Рис. 2. Эпюр речевого сигнала слова «Унтар»
 Fig. 2. Plot of the speech signal of the word «Untar»

Далее рассмотрим выделение слога из речевого сигнала, что является конечным этапом предварительной обработки.

Выделение слога из речевого сигнала. Как упоминалось, поскольку в монгольском языке слова обычно образуются добавлением суффиксов к корням слов, то, следовательно, распознавание по слогам не только улучшает способность распознавания, но и дает возможность повысить вероятность распознавания в шумной обстановке, так как шумовой сигнал между слогами также распознается нейронной сетью.

Результат 5-го этапа предварительной обработки служит входом этапа выделения слога из речевого сигнала, и вначале сигнал проходит через полупериодный выпрямитель, который описывается уравнением [6]

$$y(n) = \frac{1}{2}x(n)[1 + \text{sign}(x(n))],$$

$$\text{где } \text{sign}(x) = \begin{cases} -1, & x < 0; \\ 0, & x = 0; \\ 1, & x > 0. \end{cases}$$

На рис. 3 показаны результаты использования полупериодного выпрямителя.

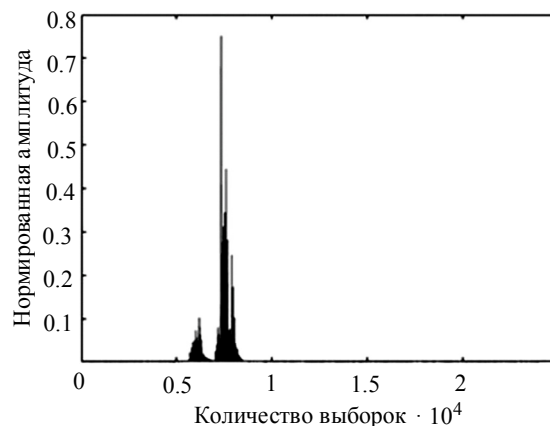


Рис. 3. Вид сигнала после полупериодного выпрямителя
 Fig. 3. Signal type after half-wave rectifier

Затем был использован нелинейный фильтр типа «envelope» для выделения огибающей сигнала [7], результат показан на рис. 4.

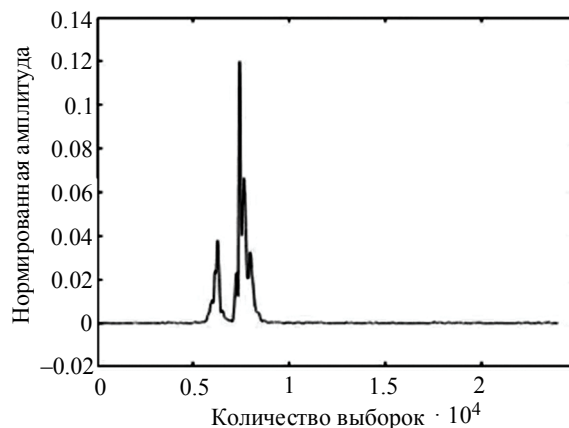


Рис. 4. Вид сигнала после выделения огибающей
 Fig. 4. Signal view after envelope selection

Благодаря выделенной огибающей речевого сигнала можно различать слоги, и в зависимости от отношения сигнал/шум – начало и конец слога в слове [8], [9] (рис. 5).

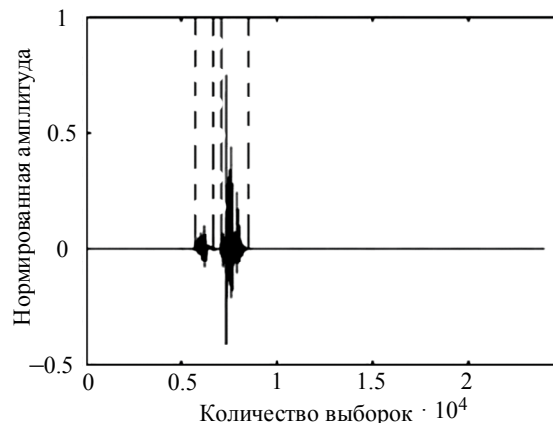


Рис. 5. Результат выделения слогов в слове «Унтар»
 Fig. 5. The result of the selection of syllables in the word «Untar»

На эпюре входного сигнала визуально можно видеть слоги в слове «Унтар», но для выделения слога нужно распознать начало и конец слога по таким параметрам речевого сигнала, как скорость пересечения нуля, кратковременная энергия и т. п. Разделы слога отмечены штриховой линией. Как видно на рис. 5, первый – «Ун» и второй – «тар» слоги слова «Унтар» были успешно разделены.

Выделение признаков речевого сигнала. После выделения слога был выполнен процесс, позволяющий выделить признаки слога речевого сигнала. Частота речевого сигнала – один из главных параметров речевых сигналов, который используется как признак при выделении слога в слове. В этом исследовании был использован мел-кепстральный коэффициент частоты (MFCC) [10]. Последовательность выделения MFCC показана на рис. 6.

Чтобы получить коэффициент MFCC, сначала, используя быстрое преобразование Фурье, преобразуем сигнал в частотную область:

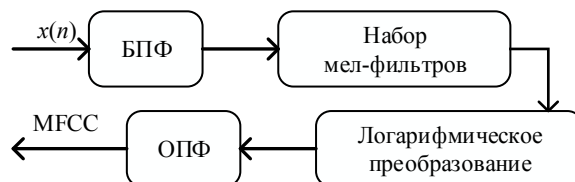


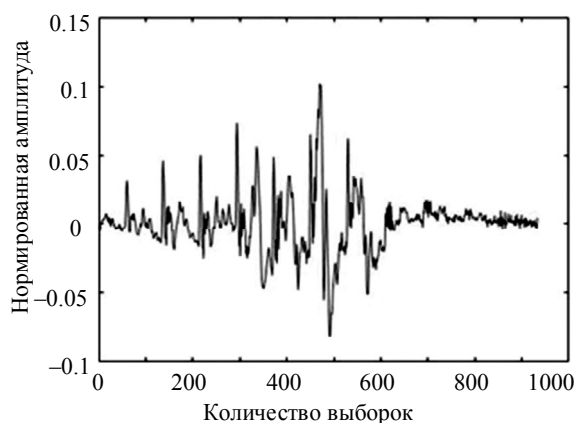
Рис. 6. Последовательность выделения MFCC речевого сигнала, где БПФ – быстрое преобразование Фурье, ОПФ – обратное преобразование Фурье

Fig. 6. Sequence of MFCC extraction of a speech signal, where БПФ is the fast Fourier transform, ОПФ is the inverse Fourier transform

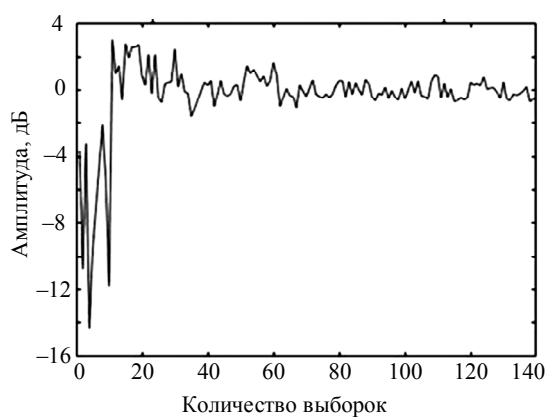
$$X(k) = \sum_{n=1}^N x(n) e^{-\frac{j2\pi nk}{N}}.$$

После этого создаем банк мел-частотных фильтров [11]:

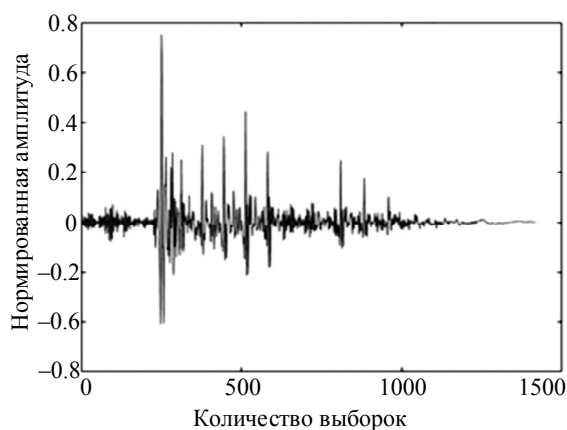
$$Y(q) = \sum_{k=1}^K |\dot{X}(k)| U_q(k),$$



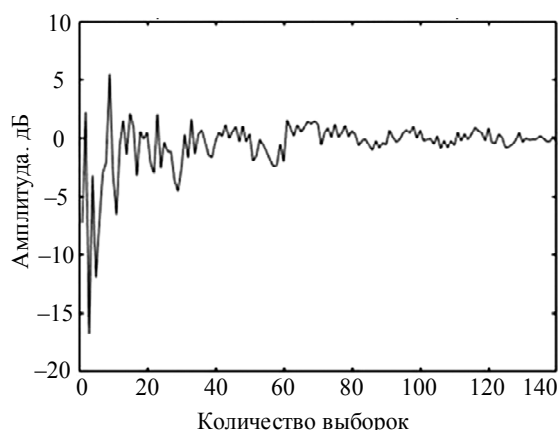
a



б



в



з

Рис. 7. Эпюры и коэффициенты MFCC в речевом сигнале слова «Унтар», разделенного по слогам «Ун» и «тар»: а – эпюра слога «Ун»; б – MFCC коэффициент слога «Ун»; в – эпюра слога «тар»; з – MFCC коэффициент слога «тар»

Fig. 7. Plots and MFCC coefficients in the speech signal of the word «Untar» divided by syllables «Un» and «tar»: а – Plot of the syllable «Un»; б – MFCC coefficient of the syllable «Un»; в – Plot of the syllable «tar»; з – MFCC coefficient of the syllable «tar»

где

$$U_q(k) = \begin{cases} 0, & k < f(q-1); \\ \frac{k - f(q-1)}{f(q) - f(q-1)}, & f(q-1) \leq k \leq f(q); \\ \frac{f(q+1) - k}{f(q+1) - f(q)}, & f(q) \leq k \leq f(q+1); \\ 0, & k > f(q+1). \end{cases}$$

Далее преобразуем сигнал в логарифмическую шкалу.

$$L(m) = \log |Y(k)|^2.$$

Наконец, завершающим этапом будет обратное преобразование Фурье [12]

$$L(m) = \log |Y(k)|^2.$$

На рис. 7 показаны эпюры и коэффициенты MFCC для слова «Унтар», разделенного по слогам «Ун» и «тар».

Процесс классификации речевого сигнала.

Существуют различные алгоритмы классификаторов – логистическая регрессия, k-ближайший сосед, наивный байесовский классификатор, дерево решений, случайный лес, метод опорных векторов и искусственная нейронная сеть, однако алгоритмы искусственной нейронной сети более успешно используются для распознавания речевого сигнала [13], [14].

Алгоритм искусственной нейронной сети – это алгоритм обучения и состоит из входных и выходных нейронов [15]. На рис. 8 показана модель одного нейрона, где $x_1, \dots, x_n$ – входы нейрона, $w_1, \dots, w_n$ – весовые коэффициенты, ФА – функция активации.

Весовые коэффициенты показывают влияние входных данных на выход. Нейроны обладают функциями активации, из которых наиболее часто используется сигмовидная функция

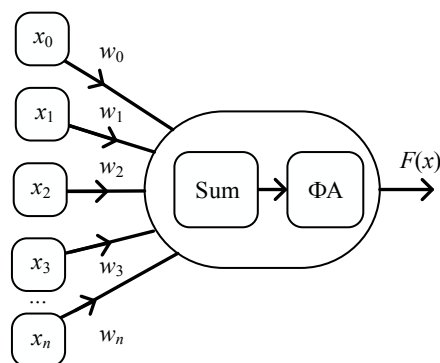


Рис. 8. Модель одного нейрона
 Fig. 8. Single neuron model

$$y = \frac{1}{1 + \exp(-\text{Sum})},$$

где $\text{Sum} = \sum_{i=1}^n x_i w_i$.

Отдельные нейроны находятся во взаимосвязи между собой и образуют искусственную нейронную сеть. В нашем случае коэффициенты MFCC, полученные после выделения признаков, подаются на вход искусственной нейронной сети.

Экспериментальная часть распознавания слова по слогам. В эксперименте использовались четыре слова монгольского языка: «ас», «унтар», «тод» и «будэг». Эти слова озвучивали 11 дикторов, мужчины и женщины разного возраста: 1 человек младше 20–30 лет, 7 человек в возрасте 30–40 лет и 3 человека – 60–70 лет. Слова разбиты следующим образом: «ас», «буд», «дэг», «г», «будэг», «ун», «тар», «унтар», «тод»; рассматривается также «шумовой сигнал». Эти слова выбраны из соображения, что они часто используются в разговорной речи и содержат основные звуки речи монгольского языка. Искусственная нейронная сеть была создана в MatLab, она имеет 112 входных и 10 выходных нейронов. На рис. 9

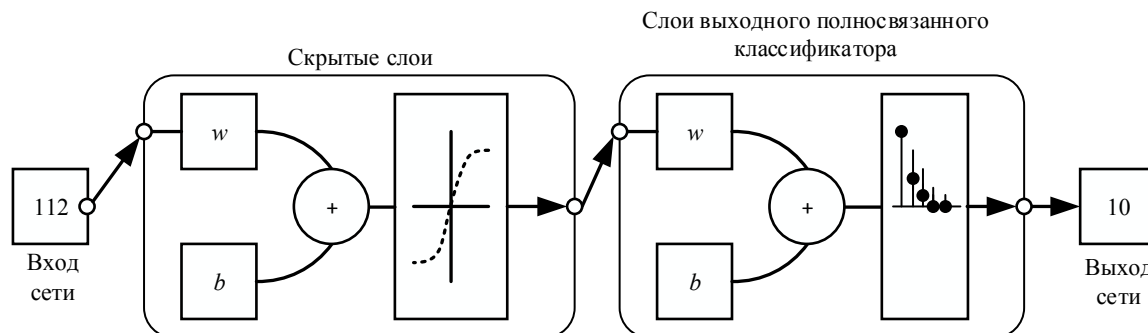


Рис. 9. Искусственная нейронная сеть, используемая в эксперименте
 Fig. 9. Artificial neural network used in the experiment

показана искусственная нейронная сеть, созданная для эксперимента, где w на диаграмме обозначает веса, а b – единицы смещения, которые являются частью отдельных нейронов.

Отдельные нейроны в скрытом слое выглядят следующим образом: 112 входных нейронов и соответствующие веса, 1 единица смещения и 10 выходных нейронов.

В процессе обучения нейронной сети использовался алгоритм обратного распространения ошибки Левенберга–Марквардта. Этот алгоритм быстрее, чем алгоритм обратного распространения ошибки, и более стабилен, чем алгоритм Гаусса–Ньютона. Этот алгоритм эффективен для обучения мелко-масштабной сети [16].

Результаты эксперимента. Процесс обучения искусственной нейронной сети показал

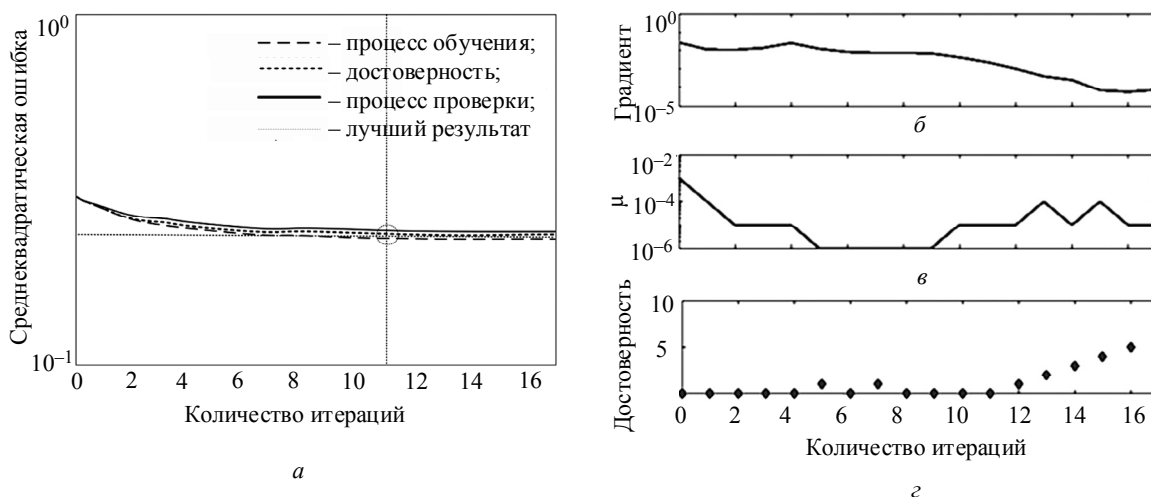


Рис. 10. Ошибки и градиенты процесса обучения сети: а – процесс обучения искусственной нейронной сети; б – градиент процесса обучения; в – параметр адаптации нейросети; г – количество проверок достоверности
Fig. 10. Errors and gradients of the network learning process: а – learning process of an artificial neural network; б – learning process gradient; в – neural network adaptation parameter; г – quantity of validation

Фактические значения	ас	146 14.1 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	2 0.2 %	2 0.2 %	0 0.0 %	0 0.0 %	95.7 %	96.1 %, 3.9 %
	буд	0 0.0 %	111 10.8 %	3 0.3 %	0 0.0 %	0 0.0 %	4 0.4 %	0 0.0 %	0 0.0 %	1 0.1 %	0 0.0 %	93.3 %, 6.7 %
	будзг	0 0.0 %	0 0.0 %	28 2.7 %	1 0.1 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	96.6 %, 3.4 %
	дзг	0 0.0 %	1 0.1 %	1 0.1 %	114 11.0 %	0 0.0 %	1 0.1 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	97.4 %, 2.6 %
	г	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	66 6.4 %	2 0.2 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	97.1 %, 2.9 %
	«шум»	0 0.0 %	1 0.1 %	0 0.0 %	0 0.0 %	1 0.1 %	108 10.5 %	1 0.1 %	1 0.1 %	0 0.0 %	0 0.0 %	96.4 %, 3.6 %
	тар	3 0.3 %	0 0.0 %	0 0.0 %	1 0.1 %	1 0.1 %	0 0.0 %	124 12.0 %	0 0.0 %	0 0.0 %	0 0.0 %	96.1 %, 3.9 %
	ун	0 0.0 %	1 0.1 %	0 0.0 %	0 0.0 %	1 0.1 %	0 0.0 %	1 0.1 %	129 12.5 %	0 0.0 %	0 0.0 %	97.7 %, 2.3 %
	унтар	0 0.0 %	0 0.0 %	0 0.0 %	1 0.1 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	19 1.8 %	0 0.0 %	95.0 %, 5.0 %
	тод	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	1 0.1 %	2 0.2 %	0 0.0 %	0 0.0 %	151 14.6 %	98.1 %, 1.9 %
		98.0 %, 2.0 %	97.4 %, 2.6 %	87.5 %, 12.5 %	97.4 %, 2.6 %	95.7 %, 2.6 %	91.5 %, 8.5 %	95.4 %, 4.6 %	99.2 %, 0.8 %	95.0 %, 5.0 %	98.7 %, 1.3 %	96.5 %, 3.5 %
		ас	буд	будзг	дзг	г	«шум»	тар	ун	унтар	тод	
		Прогнозируемые значения										

Рис. 11. Матрица ошибок для представления результатов процесса обучения искусственной нейронной сети

Fig. 11. Confusion matrix for representing the results of the artificial neural network training process

наилучшие результаты в 17-й итерации. Лучший результат проверки равен 0.23206 на 11-й итерации (рис. 10, а). На рис. 10 также показаны ошибки и градиенты процесса обучения сети. На 17-й итерации градиент $= 7.1706e^{-05}$, $\mu = 1e^{-05}$. μ (Mu) – параметр адаптации, который используется в процессе оптимизации нейронной сети Левенберга–Марквардта при вычислении обновления параметров и включается в выражение обновления веса, чтобы избежать проблемы локального минимума. Выбор μ напрямую влияет на сходимость ошибок. Диапазон μ составляет от 0 до 1. Во время обучения прогресс постоянно обновляется в окне обучения. Размер градиента и количество проверок достоверности используются для завершения обучения. Градиент станет очень маленьким, когда обучение достигнет минимума производительности. Обучение можно остановить, установив значение градиента. Количество проверок достоверности представляет собой количество последовательных итераций, на которых производительность проверки не может снизиться. Если это число достигнет 6 (значение по умолчанию для сети Левенберга–Марквардта), тренировка остановится. В этом прогоне вы можете видеть, что обучение действительно остано-

вилось из-за количества проверок достоверности. Этот критерий можно изменить.

Матрица ошибок используется для представления результатов процесса обучения искусственной нейронной сети более подробно. Матрица ошибок более четко показывает, как каждый слог пересекается с другим. На рис. 11 показана матрица ошибок процесса обучения искусственной нейронной сети, которая позволяет оценить эффективность используемого алгоритма, сопоставляя прогнозируемое и фактическое значения.

Из матрицы ошибок можно видеть, как слоги отличаются друг от друга. Она показывает, что слоги распознаны по отдельности с вероятностью от 95 до 98.1 %.

Заключение. В этом исследовании был проведен эксперимент по распознаванию речи с помощью искусственной нейронной сети, отделив слоги от речевого сигнала. Созданная искусственная нейронная сеть распознает слоги с вероятностью 96.5 %.

В дальнейшем мы будем повышать эффективность разработанной модели искусственной нейронной сети, расширив базу данных другими словами.

Список литературы

1. Speech recognition in Mongolian Language using a Neural Network with pre-processing Technique / Z. Batzorig, O. Bukhtsooj, A. G. Chensky, T. Galbaatar // Proc. of 2nd 2020 Intern. Youth Conf. on Radio Electronics, Electrical and Power Engineering (REEPE 2020). Moscow: Institute of Electrical and Electronics Engineers Inc., 2020. P. 1–5. doi: 10.1109/REEPE49198.2020.9059228.
2. A Mongolian speech recognition system based on HMM / G. Gao, Biligetu, Nabuqing, S. Zhang // Lecture Notes in Comp. Sci. 2006. Part II. Vol. 4114. P. 667–676. doi: 10.1007/978-3-540-37275-2_84.
3. Gevaert W., Tsenov G., Mladenov V. Neural Networks used for speech recognition // J. of Automatic Control. 2010. Vol. 20, Iss. 1. P. 1–7. doi: 10.2298/JAC1001001G.
4. Nandhini S., Shenbagavalli A. Voiced/Unvoiced detection using short term processing // IJCA Proc. Intern. Conf. on Innovations in Information, Embedded and Communication Systems (ICIIECS), Tamilnadu, India, Karpagam College of Engineering. 2014. Vol. 2. P. 39–43.
5. Применение искусственной нейронной сети для распознавания речевых команд на монгольском языке / Б. Зандан, А. Г. Ченский, О. Бухцоож, Т. Галбаатар // Сб. науч. тр. по материалам XIII Intern. sci. conf. Анапа: Изд-во ООО «Научно-исследовательского центра экономических и социальных процессов» в ЮФО, 2019. С. 49–54.

6. Rabiner L. R., Schafer R. W. An introduction to digital speech processing. Hanover, USA. 2007. Vol. 1, no. 1–2. P. 1–194.
7. Fernandes da Silva E. L., de Oliveira H. M. Implementation of an automatic syllabic division algorithm from speech files in portuguese language. XIX Congresso Brasileiro de Automatica CBA, Campina Grande, 2012.
8. Lipeika A., Lipeikiene J., Telksnys L. Development of isolated word speech recognition system // Informatika, Lith. Acad. Sci. 2002. Vol. 13. P. 37–46.
9. Rabiner L., Juang B.-H. Fundamentals of speech recognition. Prentice-Hall, Englewood Cliffs, N. J., 1993.
10. Lokesh, S., Devi M. R. Speech recognition system using enhanced mel frequency cepstral coefficient with windowing and framing method // Cluster Comp., 2019. Vol. 22. P. 11669–11679. doi: 10.1007/s10586-017-1447-6.
11. Memon Sh., Gregory M. A. A novel approach for MFCC feature extraction // J. of Comp. Sci. and Technol. 2011. doi: 10.1109/ICSPCS.2010.5709752. URL: https://www.researchgate.net/publication/224217606_A_novel_approach_for_MFCC_feature_extraction (дата обращения 13.04.2022).
12. Fang Z., Guoliang Z., Zhanjiang S. Comparison of different implementations of MFCC // J. Comp. Sci. Technol. 2000. Vol. 16. P. 582–589.
13. Nair R. R., Vijayan B. Voice based Gender Recognition // Intern. Research J. of Engin. and Technol. 2019. Vol. 06, Iss. 05. P. 2109–2112.

14. B. Satya Prasad. Gender classification through voice and performance analysis by using machine learning algorithms // Intern. J. of Research in Comp. Appl. and Robotics. 2019. Vol. 7, Iss. 11. P. 1–11.

15. Ala-Keturi V. Speech Recognition based on artificial Neural Networks. Helsinki University of Technology, 2004.

16. Hagan M. T., Menhaj M. B. Training feedforward networks with the Marquardt algorithm // IEEE Transactions on Neural Networks, 1994. Vol. 5, no. 6, P. 989–993. doi: 10.1109/72.329697.

Информация об авторах

Батзориг Зандан – мл. науч. сотрудник лаборатории радиоэлектроники Института физики и технологии Академии наук Монголии. 13330, Монголия, г. Улан-Батор, пр. Энхтайван, 54Б.
E-mail: zandan1984@gmail.com

Тувдэндорж Галбаатар – д-р физ.-мат. наук, академик Академии наук Монголии, вед. науч. сотрудник лаборатории радиоэлектроники Института физики и технологии Академии наук Монголии.
E-mail: tgalbaatar@mas.ac.mn

Одсүрэн Бухцоож – канд. техн. наук, доцент, науч. сотр. отделения информационных технологий Института математики и цифровых технологий Академии наук Монголии. Доцент кафедры радиотехнических приборов и антенных систем НИУ МЭИ.
E-mail: bukhtsooj@yahoo.com

Ченский Александр Геннадьевич – канд. физ.-мат. наук, зав. кафедрой радиоэлектроники и телекоммуникационных систем Иркутского национального исследовательского технического университета.
E-mail: zavmts@ex.istu.edu

References

1. Batzorig Z., Bukhtsooj O., Chensky A. G., Galbaatar T. Speech Recognition in Mongolian Language Using a Neural Network with Pre-Processing Technique // Proc. of 2nd 2020 Intern. Youth Conf. on Radio Electronics, Electrical and Power Engineering (REEPE 2020). M.: Institute of Electrical and Electronics Engineers Inc., 2020. P. 1–5. doi: 10.1109/REEPE49198.2020.9059228.

2. Gao G., Biligetu, Nabuqing, Zhang S. A Mongolian Speech Recognition System Based on HMM // Lecture Notes in Computer Science. 2006. P. II. Vol. 4114. P. 667–676. doi: 10.1007/978-3-540-37275-2_84.

3. Gevaert W., Tsenov G., Mladenov V. Neural Networks Used for Speech Recognition // J. of Automatic Control. 2010. Vol. 20, Iss. 1. P. 1–7. doi: 10.2298/JAC1001001G.

4. Nandhini S., Shenbagavalli A. Voiced/Unvoiced Detection using Short Term Processing // IJCA Proc. Intern. Conf. on Innovations in Information, Embedded and Communication Systems (ICIIECS), Tamilnadu, India, Karpagam College of Engineering. 2014. Vol. 2. P. 39–43.

5. Zandan B., CHenskij A. G., Buhcoozh O., Galbaatar T. Primenenie iskusstvennoj nejronnoj seti dlya raspoznavaniya rechevyh komand na mongol'skom yazyke // Sb. nauchn. tr. po materialam XIII Intern. sci. conf. Anapa: Izd-vo OOO «Nauchno-issledovatel'skogo centra ekonomicheskikh i social'nyh processov» v YUFO, 2019. S. 49–54. (In Russ.).

6. Rabiner L. R., Schafer R. W. An Introduction to Digital Speech Processing. Hanover, USA. 2007. Vol. 1, no. 1–2. P. 1–194.

7. Fernandes da Silva E. L., de Oliveira H. M. Implementation of an Automatic Syllabic Division Algorithm

from Speech Files in Portuguese Language. XIX Congresso Brasileiro de Automatica CBA, Campina Grande, 2012.

8. Lipeika A., Lipeikiene J., Telksnys L. Development of Isolated Word Speech Recognition System // Informatica. Lith. Acad. Sci. 2002. Vol. 13. P. 37–46.

9. Rabiner L., Juang B.-H. Fundamentals of Speech Recognition. Prentice-Hall, Englewood Cliffs, N. J., 1993.

10. Lokesh S., Devi M. R. Speech Recognition System Using Enhanced Mel Frequency Cepstral Coefficient with Windowing and Framing Method // Cluster Computer, 2019. Vol. 22. P. 11669–11679. doi: 10.1007/s10586-017-1447-6.

11. Memon Sh., Gregory M. A. A Novel Approach for MFCC Feature Extraction // J. of Comp. Sci. and Technol. 2011. doi: 10.1109/ICSPCS.2010.5709752. URL: https://www.researchgate.net/publication/224217606_A_novel_approach_for_MFCC_feature_extraction (дата обращения 13.04.2022).

12. Fang Z., Guoliang Z., Zhanjiang S. Comparison of Different Implementations of MFCC // J. Comput. Sci. Technol. 2000. Vol. 16. P. 582–589.

13. Nair R. R., Vijayan B. Voice Based Gender Recognition // Intern. Research J. of Engineering and Technology. 2019. Vol. 06, Iss. 05. P. 2109–2112.

14. B. Satya Prasad. Gender Classification through Voice and Performance Analysis by Using Machine Learning Algorithms // Intern. J. of Research in Comp. Appl. and Robotics. 2019. Vol. 7, Iss. 11. P. 1–11.

15. Veera Ala-Keturi. Speech Recognition Based on Artificial Neural Networks. In: Helsinki University of Technology, 2004.

16. Hagan M. T., Menhaj M. B. Training feedforward networks with the Marquardt algorithm // IEEE Transactions on Neural Networks, 1994. Vol. 5, no. 6. P. 989–993. doi: 10.1109/72.329697.
-

Information about the authors

Batzorig Zandan – Junior Researcher of the Laboratory of Radio Electronics of the Institute of Physics and Technology of the Academy of Sciences of Mongolia.

E-mail: zandan1984@gmail.com

Tuvdendorj Galbaatar – Dr Sci. (Phis.-Math.), Academician of the Academy of Sciences of Mongolia, Leading Researcher of the Laboratory of Radio Electronics of the Institute of Physics and Technology of the Academy of Sciences of Mongolia.

E-mail: tgalbaatar@mas.ac.mn

Odsuren Bukhtsooj – Cand. Sci. (Eng.), Researcher at the Information Technology Department of the Institute of Mathematics and Digital Technologies of the Mongolian Academy of Sciences. Associate Professor of the Department of Radio technical equipment and antenna system, MPEI, Russia.

E-mail: bukhtsooj@yahoo.com

Alexander G. Chensky – Cand. Sci. (Phis.-Math.), Department of Radio Electronics and telecommunication systems of Irkutsk National Research Technical University (IRNITU).

E-mail: zavmts@ex.istu.edu

Статья поступила в редакцию 11.04.2022; принята к публикации после рецензирования 07.05.2022; опубликована онлайн 30.06.2022.

Submitted 11.04.2022; accepted 07.05.2022; published online 30.06.2022.
