

DOI <https://doi.org/10.30898/1684-1719.2021.6.2>

УДК 004.934.2

РАСПОЗНАВАНИЕ РЕЧИ КАК ОДИН ИЗ МЕТОДОВ ОПРЕДЕЛЕНИЯ ПСИХИЧЕСКОГО СОСТОЯНИЯ ЧЕЛОВЕКА

Б. Зандан¹, А. И. Баскаков², Б. Одсурэн^{1,2}

¹ Институт Физики и Технологии Академии Наук Монголии, лаборатория радиоэлектроники, 13330, Улан-Батор, район Баянзурх, просп. Энхтайван, д. 54Б

² Московский Энергетический Институт, кафедра радиотехнических приборов, 11250, Москва, ул. Красноказарменная, 14

Статья поступила в редакцию 4 июня 2021 г.

Аннотация. В данной работе мы исследовали работу нейронной сети для распознавания речи монгольского языка и эмоционального состояния в ней. Для эксперимента мы использовали часто используемые четыре слова, которые состоят из часто употребляемых согласных и гласных. Эмоциональные состояния в речи выбраны с учетом дальнейшего исследования изменения психических расстройств человека. Для анализа мы использовали базу данных речей 12 мужчин и женщин с восемью типами эмоций. Нейронная сеть распознает слова по слогам с эффективностью до 96 процентов, а эмоции – до 80 процентов. Эти результаты показывают, что для дальнейшего анализа психических расстройств человека остается только собирать базу данных из речей пациентов соответствующих медицинских учреждений.

Ключевые слова: распознавание речи и эмоций, нейронная сеть, психическое состояние, эмоциональное состояние, монгольский язык.

Abstract. In this work, we investigated the work of a neural network for recognizing speech in the Mongolian language and the emotional state in it. For the experiment, we used four commonly used words, which are composed of frequently used consonants and vowels. Emotional states in the speech were selected taking into account further research on changes in human mental disorders. For the analysis, we used a database of

speeches of 12 men and women with eight types of emotions. Neural net recognizes words by syllables with an efficiency of up to 96 percent, and emotions - up to 80 percent. These results show that for further analysis of human mental disorders, it remains only to collect a database from the speeches of patients of the relevant medical institutions.

Keywords: speech and emotion recognition, neural network, mental state, emotional state, Mongolian language.

Введение

Распознавание речи как один из методов автоматического анализа речи позволяет анализировать психическое состояние человека, такие как одиночество [1], депрессия [2], и предсказывает развитие психоза, который сопровождают многие психические расстройства, включая, например, шизофрению и биполярное расстройство [3, 4]. Изменение психологического и психического состояния человека обнаруживается в нарушении как смысловой, так и синтаксической составляющих: например, бедность: использование простых построений и понимание только закрытых типов вопросов и путанность: нарушение связности речи [5].

Для анализа человеческой речи выбор параметров, с помощью которых мы определяем психическое состояние человека, зависит также от того, является ли речь письменной или устной. Если исследуются аудиозаписи, то должны учитываться пространственные и временные параметры речи: паузы, высота голоса и среднее время, потраченное на произнесение одного слова. Кроме того, для повышения достоверности результата должны быть вычислены самые частотные слова и словосочетания, употребляемые людьми с разным диагнозом. На основе этих параметров, используя их для выделения признаков, нейросеть выдает «диагноз» о психическом состоянии человека [2].

Поэтому для освоения методов автоматического анализа психического состояния человека на основе устной речи на монгольском языке нам необходимо было развивать нейронную сеть, которая распознает не только слова, но и слоги в слове, а также выделяют такие параметры речевого сигнала, как паузы, высота голоса и среднее время, потраченное на произнесение одного слова.

Экспериментальная часть

Для распознавания слов и слог в слове мы использовали нейронную сеть на основе метода Левенберга-Марквардта [6], после обучения нейронная сеть распознает короткие слова или слоги речи на монгольском языке с точностью 96,5 процентов. В таблице 1 приведены использованные слова и слоги, а нумерация столбцов таблице соответствует нумерации на матрице ошибок [7], используемой для оценки точности (рис.1).

Таблица 1. Слова и слоги слов на монгольском языке для обучения нейронной сети

Нумерация	1	2	3	4	5	6	7	8	9	10
Слова и слоги на монгольском языке	ас	буд	будэг	дэг	г	шуугиан	тар	ун	унтар	тод
Количество образцов для обучения	149	114	32	117	69	118	130	130	20	153

Кроме распознавания слов и слогов для анализа психического состояния человека выделяются такие параметры речевого сигнала, как паузы, высота голоса и среднее время, потраченное на произнесение одного слова. Вычисляя такие параметры речевого сигнал, как скорость речи, энергия речевого сигнала, частота основного тона, I, II и III форманты, а также I, II и III коэффициенты банка мел-фильтра [8, 9], мы распознавали восемь типов эмоций в речи человека [10, 11]:

1. Нейтральная речь
2. Спокойная речь

3. Радостная речь
4. Грустная речь
5. Сердитая речь
6. Испуганная речь
7. Речь с отвращением
8. Удивленная речь

Confusion Matrix

Output class	1	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%	2 0.2%	96.1% 3.9%
	2	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%
	3	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%
	4	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%
	5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	66 6.4%	2 0.2%	0 0.0%	0 0.0%	2 0.2%	0 0.0%	97.1% 2.9%
	6	0 0.0%	1 0.1%	0 0.0%	0 0.0%	1 0.1%	108 10.5%	0 0.0%	1 0.1%	1 0.1%	0 0.0%	96.4% 3.6%
	7	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%
	8	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%
	9	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%
	10	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% 4.3%	91.5% 8.5%	95.4% 4.6%	99.2% 0.8%	95.0% 5.0%	98.7% 1.3%
		1	2	3	4	5	6	7	8	9	10	
		Target class										

9. Рис. 1. Матрица ошибок при распознавании слов и слог в слове речи на МОНГОЛЬСКОМ ЯЗЫКЕ.

Для анализа эмоционального состояния человека с помощью речевых сигналов и обучения нейронной сети была использована база данных, состоящая из речи 12 мужчин и 12 женщин. И результат обучения и распознавания эмоционального состояния человека из его речи также приведен в виде матриц ошибок на рис. 2.

Из матрицы ошибок на рис. 2 видно, что нейронная сеть распознает эмоцию в речи человека с вероятностью 80.5 процентов. Полученные результаты дают нам

возможность сделать вывод, что мы правильно вычисляем нужные нам параметры речи для анализа и определения психического состояния человека.

Confusion Matrix

Output class	1	69 4.8%	10 0.7%	7 0.5%	6 0.4%	1 0.1%	1 0.1%	2 0.2%	0 0.0%	66.3% 33.7%
	2	15 1.0%	171 11.9%	4 0.3%	10 0.7%	0 0.0%	1 0.1%	7 0.5%	2 0.1%	81.4% 18.6%
	3	2 0.1%	0 0.0%	149 10.3%	5 0.3%	0 0.0%	4 0.3%	2 0.1%	11 0.8%	86.1% 13.9%
	4	8 0.6%	8 0.6%	8 0.6%	155 10.8%	3 0.2%	10 0.7%	15 1.0%	6 0.4%	72.8% 27.2%
	5	0 0.0%	0 0.0%	1 0.1%	4 0.3%	170 11.8%	10 0.7%	8 0.6%	3 0.2%	86.7% 13.3%
	6	0 0.0%	1 0.1%	8 0.6%	3 0.2%	3 0.2%	144 10.0%	3 0.2%	5 0.3%	86.2% 13.8%
	7	1 0.1%	1 0.1%	3 0.2%	3 0.2%	10 0.7%	8 0.6%	143 9.9%	4 0.3%	82.7% 17.3%
	8	1 0.1%	1 0.1%	12 0.8%	6 0.4%	5 0.3%	14 1.0%	7 0.5%	158 11.0%	77.5% 22.5%
			71.9% 28.1%	89.1% 10.9%	77.6% 22.4%	80.7% 19.3%	88.5% 11.5%	75.0% 25.0%	74.5% 25.5%	82.3% 17.7%
		1	2	3	4	5	6	7	8	
		Target class								

Рис. 2. Матрица ошибок при распознавании эмоции в устной речи человека: нумерация соответствует эмоциям в речи человека.

Распознавание речи с помощью нейронной сети не только позволяет анализировать психическое состояние человека, но и позволяет обнаружить и классифицировать патологии речи [12], помогает в обнаружении болезни Паркинсона [13] и идентифицировать одиночество, которое становится фактором развития диабета II типа [14] и ишемической болезни сердца [15].

Вывод и дальнейшие работы

Очевидно, что внедрение и развитие современной технологии искусственного интеллекта в различные области медицины – очень важная задача, но большинство исследований и экспериментов проведено на носителе английского или другого популярного языка. Поэтому мы начали наши исследования нейронной сети по распознаванию речи на монгольском языке с выделения необходимых для анализа речевого сигнала параметров. Нейронные сети, которые мы обучали для монгольского языка, хорошо распознают и слова, и

эмоции с эффективностью до 96 и 80 процентов, соответственно. В дальнейшем мы планируем связываться с медицинскими центрами в стране и работать именно с пациентами, которые нуждаются в той или иной поддержке: как в коррекции психического состояния, так и в лечении патологии речи.

Литература

1. Badal V.D., Graham S.A., Depp C.A., Shinkawa K., Yamada Y., Palinkas L.A., Kim H.-C., Jeste D.V., Lee E.E. Prediction of Loneliness in Older Adults Using Natural Language Processing: Exploring Sex Differences in Speech. *American Journal of Geriatric Psychiatry*. 2020. <https://doi.org/10.1016/j.jagp.2020.09.009>.
2. Al Hanai T., Ghassemi M., Glass J. Detecting Depression with Audio/Text Sequence Modeling of Interviews. *Proc. Interspeech*. 2018. P.1716-1720. <https://doi.org/10.21437/Interspeech.2018-2522>.
3. Corcoran C.M., Carrillo F., Fernández-Slezak D., Bedi G., Klim C., Javitt D.C., Bearden C.E., Cecchi G.A. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*. 2018. Vol.17. P.67-75. <https://doi.org/10.1002/wps.20491>
4. Rezaii N., Walker E., Wolff P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophr*. 2019. No.5. P.9. <https://doi.org/10.1038/s41537-019-0077-9>
5. Andreasen N.C., Grove W.M. Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophr Bull*. 1986. Vol.12. No.3. P.48-59. <https://doi.org/10.1093/schbul/12.3.348>. PMID: 3764356.
6. Sanaullah M., Chowdhury M.H. Neural network based classification of stressed speech using nonlinear spectral and cepstral features. *IEEE 12th International New Circuits and Systems Conference (NEWCAS)*. Trois-Rivieres, QC, Canada, 2014, P.33-36. <https://doi.org/10.1109/NEWCAS.2014.693397>

7. Ting K.M. Confusion Matrix. In: Sammut C., Webb G.I., editors. *Encyclopedia of Machine Learning*. Springer, Boston, MA. 2011. https://doi.org/10.1007/978-0-387-30164-8_157
8. Shaw A., Vardhan R.K., Saxena S. Emotion Recognition and Classification in Speech using Artificial Neural Networks. *International Journal of Computer Applications*. 2016. Vol.145. No. 8.
9. Gharavian D., Sheikhan M. (2010). *Emotion Recognition and Emotion Spotting Improvement Using Formant-Related Features*. Majlesi Journal of Electrical Engineering. 2010. Vol.4. No.4. <https://doi.org/https://doi.org/10.1234/mjee.v4i4.266>
10. Banse R, Scherer K.R. Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 1996. Vol.70. No.3. P.614-36. <https://doi.org/10.1037/0022-3514.70.3.614.PMID:8851745>.
11. Javid M.M., Roshan E. Speech Emotion Recognition by Using Combinations of C5.0, Neural Network (NN), and Support Vector Machines (SVM) Classification Methods. *Journal of Mathematics and Computer Science*. 2013. No.6. P.191-200. <https://doi.org/10.22436/jmcs.06.03.03>.
12. Santhana Megala S., Padmapriya R., Jayanthi B., Suganya M. Detection And Classification Of Speech Pathology Using Deep Learning. *International Journal of Scientific & Technology Research*. 2019. Vol.8. No.12. P.3045-3051.
13. Afza N., Challa M., Mungara J. Speech Processing Algorithm For Detection Of Parkinson's Disease. *International Journal of Engineering Research & Technology (IJERT)*. 2013. Vol.2. No.4. P.1798-1803.
14. Hackett R.A., Hudson J.L., Chilcot J. Loneliness and type 2 diabetes incidence: findings from the English Longitudinal Study of Ageing. *Diabetologia*. 2020. Vol.63. P.2329–2338. <https://doi.org/10.1007/s00125-020-05258-6>.

15. Valtorta N.K., Kanaan M., Gilbody S., et al., Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies. *Heart*. 2016. No.102. P.1009-1016.

Для цитирования:

Зандан Б., Баскаков А.И., Одсурэн Б. Распознавание речи как один из методов определения психического состояния человека. *Журнал радиоэлектроники* [электронный журнал]. 2021. №6. <https://doi.org/10.30898/1684-1719.2021.6.2>