Design of Russian Vocabulary Phonetic System Based on Cyclic Neural Network

Yumin Xu* , Liguo Liu

(Heilongjiang University Of Technology jixi,158100,China)

dongyuxijiaoshi06@126.com

Abstract. In order to improve the accuracy of Russian word pronunciation, this paper proposes a design of Russian word pronunciation system based on Recurrent neural network. Firstly, the traditional Russian phoneme set has been improved and designed to better express the pronunciation characteristics of words. On the basis of the new phoneme set, a Russian pronunciation dictionary containing 20000 words has been constructed. The Recurrent neural network (RNN) framework is used to implement this algorithm. This algorithm utilizes the encoder LSTM to convert Russian words into vector representations with fixed dimensions, and then converts the vectors into the target pronunciation sequence through the decoder LSTM. Through this method, the conversion process from Russian words to their corresponding pronunciations has been achieved. Finally, a fully functional Russian vocabulary pronunciation system was designed and implemented, including interactive word pronunciation and other functions. In the experiment, the system achieved significant results on the external word test set, with a word form accuracy rate of 75% and a phoneme accuracy rate of 95%. Compared to the traditional Phonetisaurus method, the performance is better. This means that the system has significant support and advantages in providing Russian pronunciation dictionary construction. This research has provided important progress for the development of Russian Speech processing, and a feasible solution for improving the accuracy and automation of Russian vocabulary pronunciation. In addition, the improved phoneme set design and the algorithm implementation based on Recurrent neural network also provide valuable reference and enlightenment for other languages' lexical phonetic systems.

Keywords: Circulatory neural network; Russian vocabulary; Phonetic system

1 Introduction

The recurrent neural network is a kind of recurrent neural network which takes sequence data as input, recursion the evolution direction of the sequence, and all nodes (circulating units) are connected in chain. The research on recurrent neural networks began in the 1980s and 1990s, and developed into one of the deep learning algorithms at the beginning of the 21st century. Among them, bidirectional recurrent neural

G. Guan et al. (eds.), *Proceedings of the 2023 3rd International Conference on Education, Information Management and Service Science (EIMSS 2023)*, Atlantis Highlights in Computer Sciences 16,
https://doi.org/10.2991/978-94-6463-264-4_62

networks and long short-term memory networks (LSTM) are common recurrent neural networks[1].

The recurrent neural network has the characteristics of memory, parameter sharing and Turing completeness, so it has certain advantages in learning the nonlinear characteristics of sequences. The recurrent neural network has applications in Natural Language Processing, NLP), such as speech recognition, language modeling, machine translation and so on, and is also used in various time series prediction. Convolutional Neural Network(CNN) is introduced to construct a cyclic neural network, which can deal with computer vision problems including sequence input. Pronunciation dictionary is an important basic resource in the research of speech information processing, which plays a key role in speech synthesis and speech recognition system. As a phonetic alphabet, Russian constantly produces new words and foreign words in the development of language, so it is inevitable that the pronunciation dictionary will not include the pronunciation of all Russian words. Grapheme to Phoneme conversion (G2P) technology can automatically annotate Russian words and their variants, effectively solve the problem of phonetic notation of Out-of-Vocalbulary (OOV), and provide support for the construction of Russian pronunciation dictionary[2-3].

2 Russian vocabulary phonetic system based on recurrent neural network

2.1 System development environment

The development of the system is based on Ubuntu operating system, using Python programming language and supported by the recurrent neural network deep learning framework. The specific development environment is as follows:

Operating system: Ubuntu 14. 04- tumd64 LTs. Development language: Python 2.7. Deep learning framework: recurrent neural network 1.0.0.

Python development platform: Qt4.8.4+PyQt4.12+SIP4.19+QScintilla 2.8+Eric 6.1.11.

2.2 System frame structure

The overall framework of Russian vocabulary phonetic system based on recurrent neural network is shown in Figure 1.

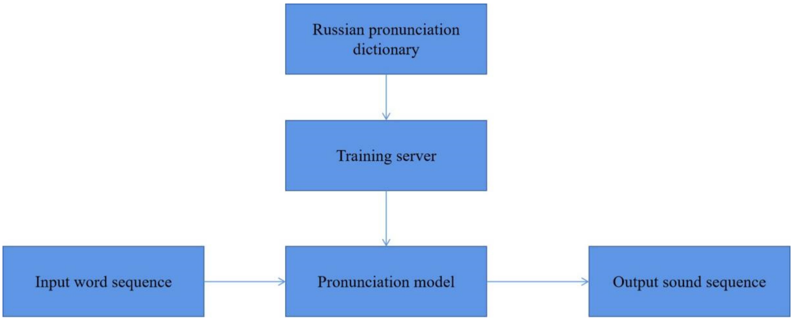

Figure1. System framework

2.3 System Function and Implementation

The design of the system is based on PyQt toolset, QtDesigner interface designer and Erie development environment, and the GUI interface is beautified by QSS(Qt Style Sheets) language. The system mainly includes model training and word phonetic function, and its implementation method is as follows:

(1) Pronunciation model training function.

The model training function takes the pronunciation dictionary as the training corpus, defines the LSTM network by calling the tf. contrib.mn interface of recurrent neural network, and calls the model_ with_ buckets method of the f. contrib. legacy _ seq2seq interface for model training. The sgd algorithm is used to optimize the parameters, which is realized by calling the GradientDescentOptimizer method of tf. train interface, and the computational complexity of gradient update is reduced by tf. nn. sampled_softmax_Iloss method[4].

(2) Word phonetic function.

The phonetic transcription function of the system is interactive. According to the loaded pronunciation model, the input words are encoded and decoded, and the word sequence is converted into pronunciation sequence. In this process, the system will call the basic_rmn_seq2seq method of the tf. contrib. legacy_ seq2seq interface. Firstly, the input sequence is converted into vector representation by encoding LSTM, and then the state of the last hidden layer of the encoder is used as input to activate decoding LSTM.

3 Experimental test

3.1 Experimental preparation

The experimental preparation of the system includes the following contents: the preparation of pronunciation dictionary corpus, the construction of experimental environment and the formulation of evaluation indicators.

(1)Dictionary corpus preparation

A corpus is a place where language materials are stored. The corpus in modern concepts refers to the original speech stored in computer memory or the corpus text annotated with linguistic information after processing. Corpus research involves the collection, storage, processing, and statistical analysis of natural language texts, with the aim of providing objective and comprehensive data to support the development of speech recognition systems through large-scale corpora. In the field of speech recognition, the most crucial step is to select suitable corpus for training the recognition model. The requirement for the corpus is to cover all speech language phenomena as much as possible, and the data should not be too sparse. At the same time, for the multi business scenarios of listening online learning, it is necessary to train multiple models, so it is crucial to design a corpus with a large vocabulary and sufficient attribute features. In the large vocabulary continuous speech recognition system, in order to train the Acoustic model with strong robustness, the collection of corpus needs to meet the following requirements. The first is to ensure that the training corpus can include as many language and speech phenomena as possible to avoid the problem of sparse Acoustic model training data. The second is to have complete phoneme coverage, which means that every smallest recognition unit in the recognition system should appear in the designed speech corpus. To ensure the accuracy of Acoustic model training, the number of occurrences of each smallest recognition unit in the recognition system in the corpus should be greater than a certain value. The third is to have a balanced phoneme, which means that the number of occurrences of each phoneme unit in the corpus should not deviate significantly from other phoneme units. A reasonable phoneme balance can effectively control the size of a speech corpus while ensuring phoneme coverage.

In the experimental stage, this paper first completed the construction of Russian pronunciation dictionary. The main sources of original corpus include Wikipedia, CMU resource database and some open source Russian corpora. The acquisition of data is realized by writing a crawler program, and the data is properly discriminated and sorted manually[5].

(2)Experimental environment

The model training and testing work of the experiment are carried out on the server. The hardware configuration of the server is: Dawning Cloud Map W760-G20 high-performance server, 16-core 7-Xeon CPU,128 GB memory and 4 x600 GB hard disk.

(3) Evaluation indicators.

The evaluation indexes to measure the word-to-sound conversion algorithm are phoneme accuracy and word shape accuracy respectively. There are generally three types of phoneme errors, namely, insertion errors, deletion errors and substitution errors. The calculation formula of phoneme accuracy is as follows:

$$\text{Correct rate of part of speeh} = \frac{\text{The nume of corrtect phonetic forms}}{\text{Total nume of word forms}} \times 100\% \tag{1}$$

4 Experimental results

In this paper, the pronunciation dictionary of 20000 words is divided into two parts, 90% as training data and 10% as test data. In the model training stage, the influence of model parameters on the system performance is observed by adjusting the layers and units of the LSTM network. In the test stage, this paper uses the method of comparative verification, and compares and tests the Phonetisaurus tools proposed by the four pronunciation model methods obtained by training. In addition, in order to measure the influence of data sources on system performance, training corpus (in-set words) and test corpus (out-of-set words) are used as test data respectively.

From the experimental results in Table 1, it can be seen that the number of layers and units of the LSTM model will have a significant impact on the system performance. When the number of layers is 3 and the number of units is 510, the performance of the system is the best. The correct rate of phonemes and word forms in the set test reaches 99.2% and 95. 8%. The correct rate of phoneme and the correct rate of word form in the out-of-set word test reached 95% and 75%, both of which were higher than the Phonetisaurus method[6-7].

Table 1. Comparison of Correct Rate of 1LSTM Model

model parameter	Correct rate of words in set/%		Correct rate of out-of-set words/%	
	Phoneme	morphology	Phoneme	morphology
LSTM layer number =2, unit number = 65.	96	80	93	65
LSTM layer number =2, unit number =510.	98	88	92	688
LSTM layer number =3, unit number = 65.	97	89	95	77
LSTM layer number =3, unit number =510.	99	96	95	75

Table 2 shows the operating efficiency comparison of four different models. The number of layers and units of the model will affect the training time, model size and decoding speed. Although the system performance is the best when the number of layers of LSTM model is 3 and the number of units is 510, the improvement of training time and the slowdown of decoding speed lead to a significant decline in system efficiency.

Table 2. Efficiency comparison of LSTM model with different parameters

model parameter	Training time /min	Model size	Reading time per word /ms
LSTM layer number =2, unit number = 65.	29	530KB	36
LSTM layer number =2, unit number =510.	156	42KB	240
LSTM layer number =3, unit number = 65.	39	730KB	40
LSTM layer number =3, unit number =510.	178	43KB	265

In order to analyze the influence of the training data scale on the system performance, this paper changes the scale of the training set, uses the LSTM network with three layers and 510 units to train the pronunciation model, and verifies the correct rate of word forms on the test set of in-set words and out-of-set words respectively. With the increasing scale of training data, the correct rate of word form is gradually improved; Under the same training scale, the phonetic correct rate of words in the set is higher than that of words outside the set[8].

4.1 Experimental analysis

The experimental results show that increasing the number of layers and units of LSTM model can improve the system performance, and when the number of network layers increases from 2 to 3, it has little influence on the model. When the number of units is increased from 64 to 510, the model size training time and system performance will be greatly improved, but the system efficiency will also be greatly reduced. When the number of layers of the LSTM model is 3 and the number of units is 510, the system performance is the best. Compared with the Phonetisaurus method, the phoneme accuracy rate is increased by 2.3 percentage points, and the word shape accuracy rate is increased by 9.5 percentage points. In addition, the scale of training corpus will also affect the system performance. With the increase of training scale, the system performance will gradually improve. But at the same time, it can be found that there is always a gap of 10% between the phonetic correctness rate of the words outside the set and the words inside the set under the condition of limited training corpus. Therefore, in order to improve the overall performance of the phonetic system, it is necessary to further expand the Russian pronunciation dictionary, improve the accuracy of the model and expand the coverage of words in the set[9-10].

5 Conclusion

Vocabulary phonetic technology can provide key support for the construction of Russian speech synthesis and speech recognition systems. This paper designs and implements a Russian vocabulary phonetic system based on Recurrent neural network. The system uses a model algorithm based on LSTM sequence to sequence. The experimental results show that the sequence to sequence model based on LSTM has achieved excellent performance in the problem of Russian word to sound conversion, and the system can be effectively applied to provide support for the construction of Russian pronunciation dictionaries. However, in the case of limited training corpus, there is a certain gap in the accuracy of the system's pronunciation of words outside the set compared to words within the set, and further improvement is needed. Therefore, in future work, it is necessary to further expand the Russian pronunciation dictionary, expand the scale of training corpus, and explore ways to improve the accuracy of the model.

Acknowledgment

Special Fund Project for Basic Scientific Research Business Expenses of Provincial Undergraduate Universities in Heilongjiang Province in 2021

A Study on the Ideological and Political Issues of Foreign Language Majors in Universities

number :2021-KYYWF-1209

References

1. Garip, Y. , & Garip, Z. . (2021). Application of artificial neural network for prediction of the cyclic oxidation behavior of electrical resistance sintered gamma-tial intermetallics. Archives of Metallurgy and Materials, 66(2), 581-591.
2. Sang, H. P. , Yoon, D. , Kim, S. , & Zong, W. G. . (2021). Deep neural network applied to joint shear strength for exterior rc beam-column joints affected by cyclic loadings. Structures, 33(1), 1819-1832.
3. Aghaei, M. H. , Baghban, M. H. , Hashemi, E. S. , & Hashemi, S. A. . (2022). Predicting effective parameters in cyclic behavior of reinforced masonry walls with shotcrete using artificial neural networks. Solid State Phenomena, 32(9), 71-78.
4. Kim, S. J. , Jeong, J. , Jang, H. W. , Yi, H. , & Lim, J. A. . (2021). Neurofiber transistors: dendritic network implementable organic neurofiber transistors with enhanced memory cyclic endurance for spatiotemporal iterative learning (adv. mater. 26/2021). Advanced Materials, 33(26)14.
5. Yang, Y. . (2021). Accurate recognition method of human body movement blurred image gait features using graph neural network. Mobile Information Systems, 2021(4), 1-11.
6. Li, Y. , & You, C. . (2021). Adaptive trading system based on lstm neural network. Journal of Physics: Conference Series, 1982(1), 012091 (5pp).
7. Tang, Z. , Liu, J. , Yu, C. , & Wang, Y. K. . (2021). Cyclic autoencoder for multimodal data alignment using custom datasets. Computer Systems Science and Engineering, 39(1), 37-54.
8. Li, W. , Wang, X. , & Feng, Q. . (2021). Final prediction of product quality in batch process based on bidirectional neural network algorithm. IOP Conference Series: Earth and Environmental Science, 692(3), 032091 (6pp).
9. Huo, J. , Sun, W. , & Dai, H. . (2021). Research on machine vision effect based on graph neural network decision. Journal of Physics: Conference Series, 1952(2), 022050 (7pp).
10. Alzahamie, Z. H. , & Abdul-Husain, H. A. . (2021). Artificial neural network for prediction of liquefaction triggering based on cpt data. Journal of Physics: Conference Series, 1973(1), 012197-.