A New Method for Semantic Consistency Verification of Aviation Radiotelephony Communication Based on LSTM-RNN

Yujun Lu+, Yihua Shi+, Guimin Jia+, Jinfeng Yang*
Tianjin Key Lab for Advanced Signal Processing
Civil Aviation University of China
Tianjin, China
jfyang@cauc.edu.cn

Abstract—**In Aviation Radiotelephony Communication (ARC), the incorrect readback between pilots and air traffic controllers has a vital effect on aircraft flight safety. To make aircraft safer in aviation, International Civil Aviation Organization (ICAO) has improved the communication standard of air traffic. However, the accidents caused by incorrect readback of ARC still happen unavoidably. To reduce the risk of incorrect readback, this paper proposes a method verifying the semantic consistency of the ARC. We firstly apply Recurrent Neural Network (RNN) and Long Short-Term memory Recurrent Neural Network (LSTM-RNN) to extract the semantic meaning of ARC and represent it with semantic vector, and then add a sigmoid layer at the output of RNN or LSTM-RNN to verify the semantic consistency. The RNN or LSTM-RNN are trained in a supervised learning method. We evaluate the proposed architecture on ARC corpus. The experimental results show that the proposed method is effective in semantic consistency verification of ARC, and LSTM-RNN outperforms the RNN in this task.**

Keywords—ARC; semantic consistency; RNN; LSTM-RNN;

I. INTRODUCTION

There are a lot of factors leading to accidents in the civil aviation transportation. Among them natural factors, mechanical failure and human factors are dominate [1]. With the development of science and technology, there has been a sharp decline in accident rate caused by natural factors and mechanical failure. However, the accident rate caused by human factors is not decreased. In the civil aviation transportation, the main human factors are operational error, maintenance failure and ARC problem. The ARC problem has a significant effect on the civil aviation safety [2]. To solve the problem of ARC, ICAO has improved air traffic communication standard. But the incorrect readbacks of ARC still exist as a result of lingering fatigue, ferocious stress, poor attention of pilots and air traffic controllers. For example, the air disaster of Tenerife in 1977 and the air crash of Urumqi in1993 were common in accident caused by incorrect readback [3]. So verifying the semantic consistency of ARC has important practical significance for the safely of civil aviation transportation. In this paper, we propose a new method to verify the semantic consistency of readbacks between pilots and air traffic controllers based on RNN.

RNN is excellent in processing arbitrary length sequences. In representing sentence semantic vector, RNN activates a word vector for each timestep in turn, which can maintain the contextual information [4]. In training, RNN is difficult to address the vanishing gradient problem and the exploding gradient problem [5]. LSTM-RNN is developed to solve the vanishing gradient problem and capture long-term dependency [6]. Due to this advantage, LSTM-RNN has been successfully applied to a variety of sequence tasks and language modeling [7]. For instance, it achieves excellent performance in speech recognition [8,9] and has been successfully applied to machine translation [10,11]. In machine translation, two LSTM-RNN models are used to represent input sentence semantic vector and generate a output sentence respectively. Similar to machine translation architecture, there are two LSTM-RNN models in information retrieval [12]. One LSTM-RNN is used to map query into a vector, and another LSTM-RNN is used to map clicked document. In this paper, we also adopt two LSTM-RNN models to represent the semantic vectors of radiotelephony communication between pilots and air traffic controllers, and then verify its semantic consistency.

In the rest of this paper, section Ⅱ describes RNN and LSTM-RNN model for semantic representation, section Ⅲ represents the learning method of RNN and LSTM-RNN, section Ⅳdiscusses the experiment results, and the last section concludes the content of this paper.

II. MODEL

A. Althgorithm procedure

To obtain semantic vector, we put the sentence pair of radiotelephony communication between pilots and air traffic controllers into RNN or LSTM-RNN model. The output of RNN or LSTM-RNN is the semantic vector utilized to verify the semantic consistency of ARC. Our algorithm procedure is as following:

+ These authors contributed equally to this work and should be considered co-first authors.
* Corresponding author.

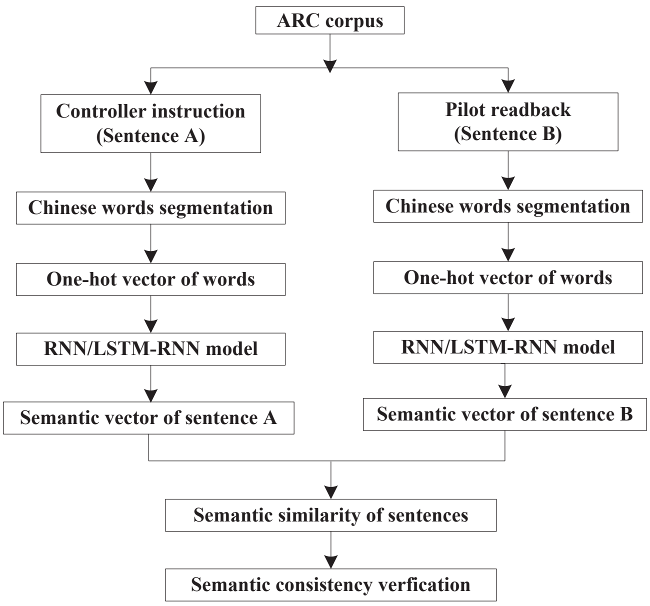

Fig.1. The algorithm procedure

In this paper, two models should be trained at the same time, one for sentence A and the other for sentence B. The training processes of two models are the same. The output vectors of models are the semantic representation of sentence pair. We adopt the cosine similarity of output vectors as the semantic similarity of sentence pair. The similarity is used to verify the semantic consistency of sentence pair.

B. RNN model

RNN is a natural feed-forward neural network. The basic architecture of RNN is show in Fig.2. We consider **X**=($x_1,x_2,\ldots,x_T$)as the input word sequence and Y=($y_1,y_2,\ldots,y_T$) as the hidden sequence. RNN's hidden layer is recurrent layer which is used to capture the contextual information of sentence [13]. RNN is sensitively to time sequence and map every word sequence to a low dimensional semantic vector. **X** is the input word sequence, $x(t)$ is the *t-th* word and $y(t)$ is the hidden activation vector in the corresponding time. In this task, $x(t)$ is coded as a one-hot vector, $y(t)$ is the semantic vector of the *t-th* word and $y(T)$ is the semantic representation of the whole sentence, when the input sequence is the last word sequence $x(T)$.

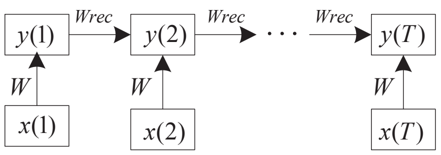

Fig.2. The basic architecture of RNN

The RNN computes the hidden activation vector by the following mathematical formulation:

$$y(t) = \tanh(Wx(t) + W_{rec}y(t-1) + b) \tag{1}$$

Where $\boldsymbol{W}$ and $\boldsymbol{W_{rec}}$ are the input weight matrix and recurrent weight matrix. b is the bias. $tanh(\cdot)$ is the activation function of hidden layer. we set bias as zero in our work.

C. LSTM-RNN model

Compared with RNN, LSTM-RNN adds output gate, forget gate, input gate and memory cell which is used to store important information over long time duration [14]. The output gate, forget gate and input gate of LSTM-RNN are respectively used to take itself offline, delete its information and neglect incoming activations [15]. The architecture of LSTM-RNN is show in Fig.3. LSTM-RNN not only is affected by the input of previous layer and the pervious time state of current layer, as well as the internal state of the LSTM-RNN. At the same time, LSTM-RNN is superior to RNN in avoiding the vanishing gradient problem.

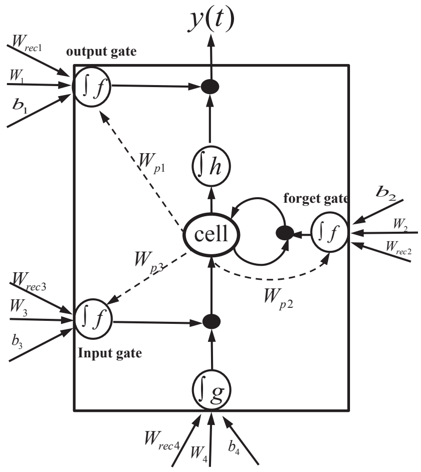

Fig.3. The architecture of LSTM-RNN

We respectively set $o(t)$, $f(t)$, $i(t)$, $c(t)$ as output gate, forget gate, input gate and cell state vector. The LSTM-RNN is computed as follows:

$$i(t) = \sigma(W_3x(t) + W_{rec3}y(t-1) + W_{p3}c(t-1)) \tag{2}$$

$$f(t) = \sigma(W_2x(t) + W_{rec2}y(t-1) + W_{p2}c(t-1)) \tag{3}$$

$$o(t) = \sigma(W_1x(t) + W_{rec1}y(t-1) + W_{p1}c(t-1)) \tag{4}$$

$$l(t) = \tanh(W_4x(t) + W_{rec4}y(t-1)) \tag{5}$$

$$c(t) = f(t)c(t-1) + i(t)l(t) \tag{6}$$

$$y(t) = \tanh(c(t))o(t) \tag{7}$$

where W_i and W_{reci} (i=1,2,3,4) are input connections and recurrent connections of output gate, forget gate, input gate and cell. W_{pi} (i=1,2,3) are peephole connections. $\sigma(.)$ is the sigmoid function. We use tanh(.) and $\sigma(.)$ as activation function. Same as RNN, $y(T)$ is the semantic representation of the whole sentence, when the input sequence is the last sequence $x(T)$.

III. LEARNING METHOD

In this work, we should make the semantic vectors of consistent sentence pairs as close as possible. To measure the semantic vectors' similarity, we compute their cosine similarity. The mathematical formula of cosine similarity is:

$$R = \frac{y(A)^T y(B)}{\|y(A)\| \cdot \|y(A)\|} \tag{8}$$

where $y(A)$ and $y(B)$ are the semantic vectors of sentence A and sentence B.

We now describe how to train the model. Our corpus is made by ourselves, which is labelled. So we utilize supervised learning method to evaluate our model parameters. In this paper, if the sentence pair is consistent, R should close to 1 as much as possible, otherwise R should close to -1 small as much as possible. We train the RNN and LSTM-RNN model by minimizing the cross-entropy error. The error is defined as :

$$L(\Lambda) = \sum_{n=1}^{N} C\log(R) + (1-C)\log(1-R) \tag{9}$$

where Λ denotes the parameters of the model and R is cosine similarity. C is labelled number, if the sentence pair is consistent, C is 1, otherwise C is 0.

We need to compute the gradient of the cross-entropy error, and obtain the minimum of the gradient with Back Propagation Though Time (BPTT) to estimate the model parameters [16]. The model parameters are updated in the process of training. In updating parameters, we adopt Nesterov method to accelerate the convergence, which is similar to standard momentum method [17]. The parameter Λ is updated at epoch k as follow equation:

$$\Delta\Lambda_k = \mu\Delta\Lambda_{k-1} - \varepsilon\nabla L(\Lambda_{k-1} + \mu\Delta\Lambda_{k-1}) \tag{10}$$

where $\nabla L(\cdot)$ is the gradient of the cross-entropy error in Eq.9, ε is the learning rate and μ is the momentum parameter.

RNN have two obviously issues, which are the vanishing and the exploding gradient problem. In section 2, we know that LSTM-RNN model can address the vanishing gradient problem by itself. To solve the exploding gradient problem, we use gradient renormalization method in the training process.

To verify the semantic consistency of sentence pairs, we add a sigmoid layer at the output of RNN and LSTM-RNN model. The input of sigmoid layer is the cosine similarity of sentence pair. If the semantic of sentence pair is consistent, the output value of sigmoid layer is 1, otherwise the output value is 0.

IV. EXPERIMENTS RESULTS

In this section, we introduce the process of our experiments. The process is divided into 3 parts.

A. ARC Corpus

We build the corpus using the recordings of radiotelephony communication between pilots and air traffic controllers. This recording are rewrote with text format by professional air traffic controllers. The sentence pairs of readback communication between pilots and air traffic are picked up, which constitute the corpus. Those sentence pairs are annotated with consistent (positive sample) and non-consistent (negative sample). In this experiment, total samples consist of 800 sentence pairs which include 500 positive samples and 300 negative samples. The total samples are split into 640 train samples and 160 test samples. The input of RNN or LSTM-RNN model is one-hot vectors of words, hence we need to segment words of sentence pairs. The example of segment words is shown in table 1.

TABLE I. EXAMPLE OF REPETITIVE SENTENCE PAIR AND SEGMENTING WORDS OF SENTENCE

Examples	
Readback sentence pair	A:国航 2763 左转 160 建立 18 左盲降 A:CA2763 turn left 160 clear for ILS approach r/w 18L established. B:左转 160 建立 18 左盲降国航 2763 B：turn left 160 clear for ILS approach r/w 18L established CA2763
Segmenting words of sentence	A: 国航/2763/左转/160/建立/18 左/盲降 B: 左转/160/建立/18 左/盲降/国航/2763

B. Glossary

In this paper, the work is related to the civil aviation transportation. we need to make a professional glossary based on the communication form in this field. The size of our glossary is 547. So the size of one-hot vector is also 547. Due to the low dimension of the one-hot vector, we avoid reducing the dimensionality of the one-hot vector, which is different from other methods.

C. Semantic consistency verification analysis

We obtain one-hot vectors of whole sentence pairs' words using glossary. To represent sentence pair with semantic vectors, the one-hot vectors of sentence pair are inputted to RNN or LSTM-RNN model. Then we utilize sigmoid layer verifying semantic consistency of sentence pair.

In our experiment, we use test accuracy and MSE evaluating the performance of RNN and LSTM-RNN model. Test accuracy is defined in Eq.11. We randomly train and test our corpus 30 times. The results are shown in Fig.3. The average test accuracy and MSE of 30 tests are presented in table 2. Average test accuracy is a measure of our models' performance and MSE gives the stability of test accuracy. The values of the error function during training for RNN and LSTM-RNN are shown in Fig.4.

$$\text{test accuracy}=\frac{\text{numbers of test samples correctly}}{\text{total numbers of test samples}} \quad (11)$$

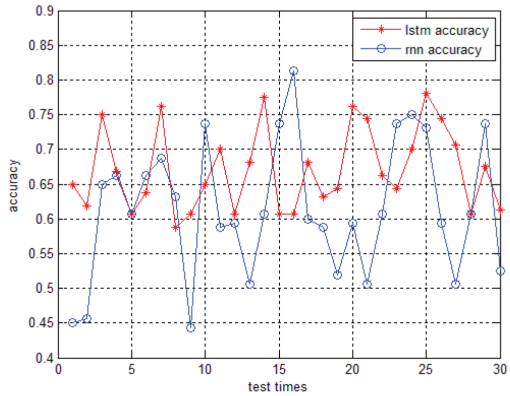

Fig.3.Test accuracy of RNN and LSTM-RNN

TABLE II. THE AVERAGE TEST A CCURACY AND MSE OF RNN AND LSTM-RNN

	average test accuracy	MSE
LSTM-RNN	0.6702	0.0586
RNN	0.6144	0.0966

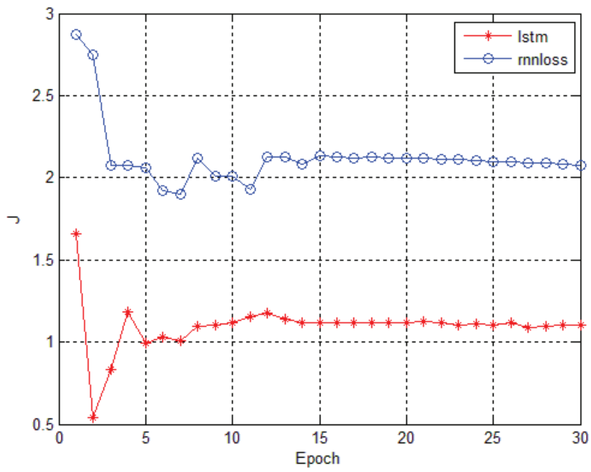

Fig.4. the training cost of RNN and LSTM-RMM

Form Fig.3, Fig.4 and table2, we can see that: (1) The test accuracy of LSTM-RNN is higher than RNN, and the MSE of LSTM-RNN is smaller than RNN. (2) The optimization of error function based on LSTM-RNN is more effective than RNN. So we can conclude that LSTM-RNN is practical in verifying semantic consistency of ARC and LSTM-RNN performs better than RNN in this task.

V. CONCLUSIONS

In the civil aviation transportation, accidents are caused by many factors. ARC problem is one of those factors. The incorrect readback between pilots and air traffic controllers is a common ARC problem. Although air traffic communication standard has been improved by ICAO, the incorrect readbacks still exist and have an effect on aircraft flight safety. To solve this problem, this paper proposes a method verifying the semantic consistency of ARC based on RNN. In this paper, we apply RNN and LSTM-RNN to represent the sentence vectors of radiotelephony communication between pilots and air traffic controllers, and use sigmoid function to verify the semantic consistency of radiotelephony communication. Finally we respectively test the performance of RNN and LSTM-RNN on ARC corpus. The results of experiment prove that LSTM-RNN is feasible to verify the semantic consistency of ARC and the performance of LSTM-RNN outperforms RNN.

ACKNOWLEDGMENT

This work is jointly supported by Joint Foundation (No.U1433120), National Science Foundation for Young Scientists of China (No. 61502498) and National Natural Science Foundation of China (No.61379102).

REFERENCES

[1] Luo, "Categorization Norm for Human Factor Accidents and Accidental Signs and the Statistics of China Civil Aviation in Recent Twelve Years," China Safety Science Journal,2002, 12(5), pp.1-12.

[2] A. C. Boschen, R. K. Johns, "Aviation Language Problem:Improving Pilot-Controller Communication," International Professional Communication Conference, 2004, pp. 291-294.

[3] WJ Pan, L Wu, HQ Chen, XL Luo, "Mistakes in Radio Communications Analysis Based on Statistic," Science and Technology of West China, 2008, 7(30), pp.1-3.

[4] I. Sutskever, O. Vinyals, and Q.V. Le, "Sequence to sequence learning with neural networks," In Proceedings of Advances in Neural Information Processing Systems, 2014, pp. 3104-3112.

[5] R. Pascanu, T. Mikolov and Y. Bengio, "On the difficulty of training recurrent neural networks," Computer Science, 2013, 52(3), pp.337-345.

[6] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Computation, 1997.

[7] T. Mikolov, "Statistical Language Models based on Netural Networks," ph.D.thesis, Brno University of Tecnology, 2012.

[8] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[9] A. Graves, A. Mohamed, and G. Hinton. "Speech recognition with deep recurrent neural networks," In Proc. of ICASSP, pp. 6645-6649, Vancouver, Canada, May 2013. IEEE.

[10] I. Sutskever, O. Vinyals, and Q.V. Le, "Sequence to sequence learning with neural networks," In Proceedings of Advances in Neural Information Processing Systems, 2014, pp. 3104-3112.

[11] H. Le, A. Allauzen, and F. Yvon, "Continuous Space Translation Models with Neural Networks," In Proc. HLT-NAACL, 2012, pp. 39-48.

[12] H. Palangi, L. Deng, Y. Shen, J. Gao, X. He, J. Chen, X. Song, R. Ward, "Deep Sentence Embedding Using the Long Short Term Memory Network: Analysis and Application to Information Retrieval," arxiv, 2015, 24(4) : 694-707.

[13] T. Mikolov, S. Kombrink, L. Burget, J. Cernocky, and S. Khudanpur, "Extensions of recurrent neural network based language model," in Proc. ICASSP, 2011, pp. 5528-5531.

[14] F.A. Gers, N.N. Schraudolph, and J. Schmidhuber, "Learning precise timing with lstm recurrent networks," Journal of Machine Learning Research, 2003, 3, pp. 115–143.

[15] A. Graves, N. Beringer, and J. Schmidhuber, “Rapid retraining on speech data with lstm recurrent networks,” Technical Report IDSIA-09-05, IDSIA, 2005.

[16] I. Sutskever, “Training Recurrent Neural Networks,” Ph.D. thesis, University of Toronto, 2013.

[17] I. Sutskever, J. Martens, G. Dahl, , and G. Hinton, “On the importance of initialization and momentum in deep learning,” In ICML (3)’13, 2013, pp. 1139–1147.