Predicting Ultrafast Nonlinear Dynamics in Fibre Optics with a Neural Network - Final Report

Zhuoru Zeng
School of Science and Engineering
The Chinese University of Hong Kong, Shenzhen
Shenzhen, China
119010417@link.cuhk.edu.cn

***Abstract*—Development of machine learning has provided researchers with new tools to study an important field of physics, the Nonlinear Schrödinger Equation System. Our project will focus on predicting the propagation of ultrashort pulses in optic fibre using Recursive Neural Networks. Our work will start by examining and tuning the existing models proposed by other scholars with our first-hand data. In the later stage of our research, we will try to make use of the ideas in Deep Complex Network and enable our Recursive Neural Networks to learn from pulses in complex-valued representation.**

***Index Terms*—Recurrent Neural Network, Fibre Optics, Signal Propagation Prediction, Long Short-Term Memory, Deep Complex Network**

I. Introduction

With the development of machine learning and its successful application in different fields, there has also been an increasing interest in applying the tools of machine learning in studying dynamical complex systems that evolve overtime [1]. These systems display strong sensitivity to small variations of the governing parameters, and using the traditional numerical methods to understand and predict these dynamic is challenging [1].

In the specific scenario of nonlinear pulse propagation in optical-fibre, many machine learning methods have been applied in various ways to analyse the spectrum or temporal intensity profile at fibre output [1]. However, the existing applications have been restricted to genetic algorithms or to feedforward neural network architectures and are limited to determine the correspondence between a given input and some single output parameter [1].

Inspired by these studies, predicting the propagation of ultrashort pulses in optical fibre is chosen to be the focus of the work of L. Salmela et al [1]. My work will start by replicating the results of their work and try to apply their model to our first-hand data with hyper-parameters tuned.

The optical fibres is a typical nonlinear Schrödinger equation (NLSE) system while the studying of its nonlinear dynamics has arouse common interests [1]. The challenge faced during the study of propagation in an NLSE systems is that these systems depends sensitively on both the input pulse and the fibre characteristics, the design and analysis of experiments require extensive numerical simulations based on the numerical integration of the NLSE [1]. The numerical simulation is computationally demanding and creates server bottleneck in using these techniques to design or optimize experiments in real time [1].

The solution proposed by Salmela's team is to use machine learning to predict the complex nonlinear propagation in optic fibre with a recurrent neural network (RNN) instead of calculating numerical solution of a governing propagation model [1]. In their experiment, a long short-term memory (LSTM) RNN is used to reproduce the complex nonlinear dynamics of ultrashort pulse evolution in optical fibre governed by an NLSE system [1]. In this system the temporal intensity profile at particular distance is depend on the intensity of earlier distance. Based on this observation, RNN architecture is chosen in building neural network model because it is widely proven to be performing well in modeling sequential dependencies [1]. The model's prediction results are compared with experimental measurements to validate accuracy.

Salmela team's work has demonstrated that RNN with LSTM can learn the complex dynamics with the nonlinear propagation of short pulses in optic fibres and the RNN model's prediction results has excellent correspondence with NLSE simulations in various experiments [1]. However, there is one limitation of their work. The pulses are represented by their temporal or spectral intensity, some information about the waves are lost in this form of representation comparing to the representation in frequency domain or complex time domain. The work on Deep Complex Network [2] provides us with an inspiration to help extend the predictions of pulses in optic fibre into complex field (amplitude and phase).

In the work Deep Complex Network, C. Trabelsi et al. explored the possibility of complex-valued deep neural networks and built up key atomic components for them [2]. M. Wolter et al. present a complex gated recurrent neural network cell structure in their study [3]. In the later state of our work, we will try to make use of the inspirations from Deep Complex Network and Complex Gated Recurrent Neural Network in predicting pulse propagation in optic fibre in complex field [2], [3]. The complex-valued components and complex gated RNN. has been proven to be useful in several experiments, namely the automatic music transcription, speech spectrum prediction, and Human motion prediction, which have similarities to our work in pulse propagation prediction. So, the use of complex neural network is a promising direction for our research topic.

II. METHODOLOGY

A. Predicting Propagating Signal Using Recurrent Neural Network

In the work of L. Salmela et al. [1], a RNN architecture is used. The neural network consists of an input layer, an LSTM recurrent layer, two hidden dense layers and the output layer. A graphical demonstration of the architecture is shown below.

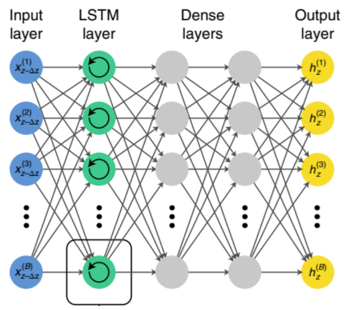

Fig. 1: Demonstration of the RNN model [1].

Training data are generated by performing NLSE numerical simulation of pulses propagation in optic fibre. Pulse intensity profile (time intensity or spectral intensity) are used as input of the model solely. The neural network uses previous intensity profiles in the evolution to yield the subsequent pulse profiles.

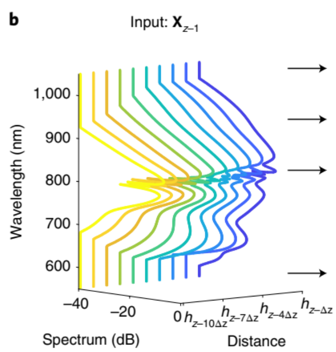

Fig. 2: Use previous intensity profile to predict subsequent pulses [1].

NLSE simulations and RNN prediction results are compared to evaluate the prediction accuracy (level of agreement between the two). Mean Squared Error (MSE) and Mean Absolute Error (MAE) are used to evaluate the difference between simulation and prediction results.

In our work, we will use the the model of L. Salmela et al. to fit on data generated by ourselves [1]. We will tune the hyper-parameters and tweak the model to try to improve the model's performance on our own data.

B. Improving Recurrent Neural Network Using Deep Complex Network Components

The limitation of Salmela team's work [1] is that the pulses are presented with temporal intensity which preserves less information comparing to the complex time domain or

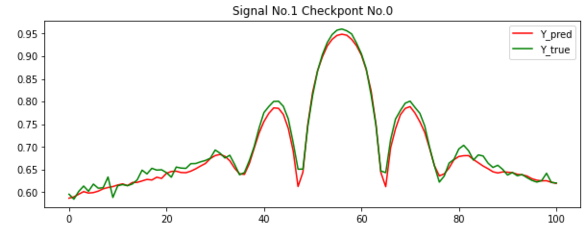

Fig. 3: Compare NLSE simulation and RNN prediction results.

frequency domain representation. With the complex waveform, we should in principle have the full information about the pulses including spectra, power, and phase. If we can use the complex representation of pulses as the input of the model, the model can be expected to be used in broader scenarios and areas.

Motivated by the limitations of the work of L. Salmela et al. [1] and inspired by the work of the study in deep complex networks presented by C. Trabelsi et al. [2], we find new directions in which we can further improve the pulse propagation prediction model by adding the components for complex-valued deep neural networks. Ideas mentioned by C. Trabelsi et al. [2], including representation of complex numbers, complex-valued activation function, complex weight initialization, will be used to improve the pulse propagation prediction model so that it can take in complex waveform as inputs and learn from these complex-valued data. M. Wolter et al. presents a complex gated recurrent neural networks structure, and implemented cgRNN units in their paper [3]. Their idea [3] of combining a gated recurrent network structure and the complex representation in sequence data prediction problem would be used to build up our model. A detailed illustration of a complex gated recurrent unit [3] is presented in the following sub sections.

1) Basic Complex RNN Formulation: The basic operations in a complex RNN is defined as follows in the study of M. Wolter et al. [3]:

$$z_t = Wh_{t-1} + Vx_t + b \tag{1}$$

$$h_t = f_a(z_t) \tag{2}$$

x_t and h_t represent the input and hidden unit vectors at time t. f_a is a point-wise non-linear activation function, and W and V are the hidden and input state transition matrices respectively. In the complex network, $x_t \in \mathbb{C}^{n_x \times 1}$, $h_t \in \mathbb{C}^{n_h \times 1}$, $W \in \mathbb{C}^{n_h \times n_h}$, $V \in \mathbb{C}^{n_h \times n_x}$ and $b \in \mathbb{C}^{n_h \times 1}$, n_x and n_h are the dimensionalities of the input and hidden states respectively.

2) Complex Non-linear Activation Function: A typical activation for deep learning is the rectified linear unit (ReLU). An extension of ReLU function from the real domain into the complex domain is proposed by M. Arjovsky et. al, the modReLU [5]:

$$f_{\text{modReLU}}(Z) = \text{ReLU}(|z|+b)e^{-i\cdot\theta_z} = \text{ReLU}(|Z|+b)\frac{z}{|z|} \tag{3}$$

modReLU activation function would be used in the implementation of the complex gated RNN cell.

TABLE I: Data Set Specification

Index	Data Set Setting	Data Set Description
1	EOcomb_space25GHz_Vratio8_bis0d24_NLSE _time_512_Power	Time domain power that describes the pulse evolvement over the whole nonlinear link
2	EOcomb_space25GHz_Vratio8_bis0d24_NALM _NLSE_time_512_power	Similar to data set 1, but the input pulse have different shape formats from data set 1
3	EOcomb_space25GHz_Vratio8_bis0d24_NALM _NLSE_time_512_field	Same representation as data set 2 except that the data is complex time domain samples

TABLE II: Experiment Result on Dataset 1

Experiment Index	Optimizer	Window Size	learning Rate	Epochs	MSE (test)	MAE (test)
1_1_0	RMSprop	10	1e-4	5	8.2850e-05	0.0069
1_1_1	RMSprop	10	1e-4	10	6.7079e-05	0.0063
1_1_2	RMSprop	10	1e-4	20	5.3493e-05	0.0055
1_1_3(Early Stopping)	RMSprop	10	1e-4	29	5.9736e-05	0.0060
1_3_0	Adam	10	1e-4	10	4.7846e-05	0.0050
1_3_1	SGD	10	1e-4	10	0.0011	0.0196
1_5_0	RMSprop	15	1e-4	10	6.4968e-05	0.0060

TABLE III: Experiment Result on Dataset 2

Experiment Index	Optimizer	Window Size	learning Rate	Epochs	MSE (test)	MAE (test)
2_1_0	RMSprop	10	1e-4	5	0.0015	0.0273
2_1_1	RMSprop	10	1e-4	10	0.0011	0.0240
2_1_2	RMSprop	10	1e-4	20	5.1074e-04	0.0160
2_1_3(Early Stopping)	RMSprop	10	1e-4	100	2.3636e-04	0.0108
2_3_0	Adam	10	1e-4	10	3.1046e-04	0.0125
2_3_1	SGD	10	1e-4	10	0.0038	0.0465
2_5_0	RMSprop	15	1e-4	10	0.0011	0.0233

3) Complex-Valued Gating Units: The complex gated RNN proposed by Wolter et al. [3] use a gated structure similar to the Gated Recurrent Unit [4]. The structure of a complex gated RNN cell is demonstrated in the figure below.

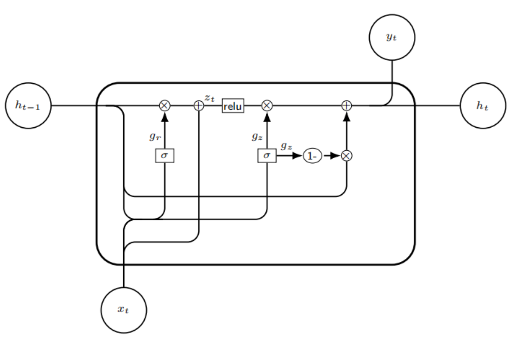

Fig. 4: Complex Gated Recurrent Neural Network Cell

The complex gate is applied as an element-wise product, i.e. $g \odot h = g \odot |h| \, e^{i\theta_h}$, and such operation results in an element-wise scaling of the hidden state's magnitude. The complex gated RNN cell is implemented as shown below:

$$z_t = W(g_r \odot h_{t-1}) + V x_t + b \tag{4}$$

$$h_t = g_z \odot f_a(z_t) + (1 - g_z) \odot h_{t-1} \tag{5}$$

h_t denotes the hidden state of the current cell, f_a denotes the activation function which is the modReLU defined in previous subsection. g_r and g_z represent reset and update gates respectively and are defined with corresponding subscripts r and z as

$$g_r = f_g(z_r), z_r = W_r h + V_r x_t + b_r \tag{6}$$

$$g_z = f_g(z_z), z_z = W_z h + V_z x_t + b_z \tag{7}$$

f_g denotes the gate activation function, $W_r \in \mathbb{C}^{n_h \times n_h}$ and $W_z \in \mathbb{C}^{n_h \times n_h}$ denotes state to state transition matrix, $V_r \in \mathbb{C}^{n_h \times n_i}$ and $V_z \in \mathbb{C}^{n_h \times n_i}$ denotes the input to state transition matrix, $b_r \in \mathbb{C}^{n_h}$ and $b_z \in \mathbb{C}^{n_h}$ denotes the biases. f_g is a complex non-linear gate activation function named modSigmoid, and is defined as below:

$$f_{\text{mod sigmoid}} = \sigma(\alpha \Re(z) + \beta \Im(z)), \alpha, \beta \in [0, 1] \tag{8}$$

III. Experimentation

A. Training Data

Three data set is introduced in the project to train and evaluate the performance of our model. Data set specification is shown in the table 1 above.

B. Improving L. Salmela et al. LSTM Model

Different hyper parameters are selected to tune the performance of our model, the experiment results are shown in table 2 and table 3 above. For experiment with index 1_1_3 and 2_1_3, the early stopping callback is used, with a maximum epoch set to 100, the early stopping criteria is that validation loss stop decreasing for more than 3 epochs. Mean Square Error (MSE) and Mean Absolute Error (MAE) is used as the evaluation criteria to the prediction performance.

For optimizer selection, both Adam and RMSprop performs relatively good, while SGD perform poor in both data set 1 and data set 2. With regard to training epochs, for data set 1, 30 epochs may be a good choice and exceeding 30 epochs may lead to overfitting. For data set 2, the validation loss continues to decrease until 100 epochs, more experiments need to be done to find the best training epochs.

An example prediction v.s. truth value graph is presented in figure 5.

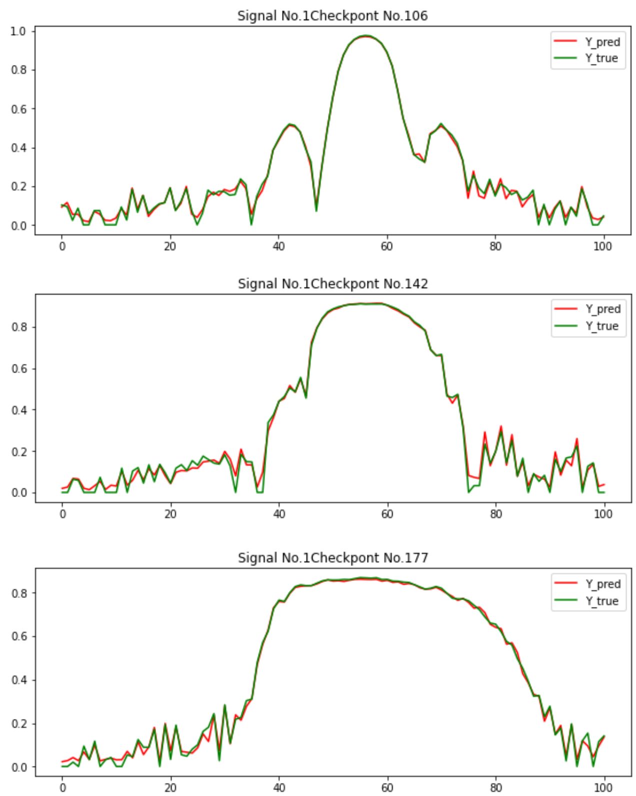

Fig. 5: Prediction Samples in experiment 2_1_2

C. Complex Gated Recurrent Unit Model

Complex representation data is used as the input data into the Complex Gated Recurrent Unit (CGRU) model. A primitive CGRU model composed of 1 layer of 100 CGRU cells is implemented. An example prediction v.s. truth value graph is presented in figure 6.

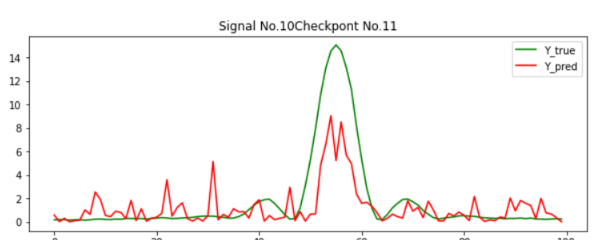

Fig. 6: Compare NLSE simulation and CGRU model prediction results.

As we can see from the figure 6, the prediction accuracy of CGRU model is relatively poor especially in comparison to previous LSTM model. Possible reasons for the poor performance of CGRU model are: First, Gated Recurrent Networks has 1 gate less than LSTM network. The simpler structure may indicates a less satisfactory performance. Second, the complex operation in CGRU model may leads to some potential information losses. Third, there may be not enough data samples for the training data set, and the model failed to learn the patterns properly. Fourth, further hyper-parameter tuning and model structure modification may be needed.

IV. Conclusion

Experiments in this project has proven that Recurrent Neural Network is a feasible solution for predicting pulse propagation in optic fibre. LSTM model is a satisfactory solution for pulse propagation prediction when amplitude or power is used as the input data. The primitive CGRU model proposed in this project is less competitive in terms of prediction accuracy, but has pointed out a feasible and promising direction of future study using complex representation of pulses.

Acknowledgement

I sincerely appreciate the kind guidance and support from my supervisor Prof. Luo Yuan who has provided me with a clear direction of researching, a lot of suggestions, and most importantly the encouragement and motivation to move forward.

References

[1] L. Salmela, N. Tsipinakis, A. Foi, et al., “Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network,” Nat Mach Intell 3, 344–354 (2021).

[2] C. Trabelsi, O. Bilaniuk, Y. Zhang, et al., “Deep Complex Networks,” Presented at ICLR 2018 Conference. [Online]. Available: https://arxiv.org/abs/1705.09792

[3] M. Wolter and A. Yao, “Gated Complex Recurrent Neural Networks”, CoRR, abs/1806.08267, 2018. [Online]. Available: http://arxiv.org/abs/1806.08267

[4] K. Cho, B. van Merriënboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. “Learning phrase representations using RNN encoder–decoder for statistical machine translation”. In EMNLP, 2014. [Online]. Available: https://arxiv.org/abs/1409.1259

[5] M. Arjovsky, A. Shah, and Y. Bengio, “Unitary evolution recurrent neural networks”.In ICML, 2016. [Online]. Available: http://arxiv.org/abs/1511.06464