ASEAN Engineering Journal

Full Paper

IMPROVING A DEEP NEURAL NETWORK GENERATIVE-BASED CHATBOT MODEL

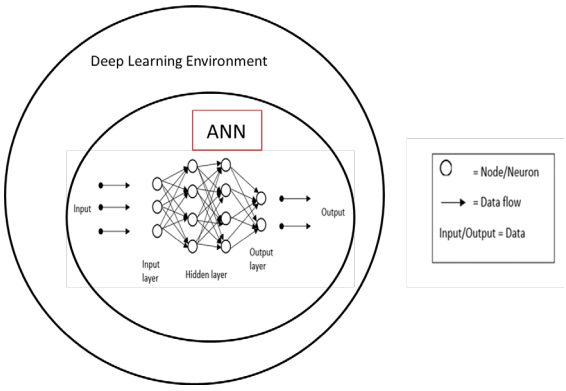
Wan Solehah Wan Ahmad*, Mohamad Nazim Jambli

Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 14300, Kota Samarahan, Sarawak, Malaysia

Article history
Received
01 July 2023
Received in revised form
25 September2023
Accepted
14 November 2023
Published online
31 May 2024

*Corresponding author
19020167@siswa.unimas.my

Graphical abstract



Abstract

A chatbot is an application that is developed in the field of machine learning, which has become a hot topic of research in recent years. The majority of today's chatbots integrate the Artificial Neural Network (ANN) approach with a Deep Learning environment, which results in a new generation chatbot known as a Generative-Based Chatbot. The current chatbot application mostly fails to recognize the optimum capacity of the network environment due to its complex nature resulting in low accuracy and loss rate. In this paper, we aim to conduct an experiment in evaluating the performance of chatbot model when manipulating the selected hyperparameters that can greatly contribute to the well-performed model without modifying any major structures and algorithms in the model. The experiment involves training two models, which are the Attentive Sequence-to-Sequence model (baseline model), and Attentive Seq2Sequence with Hyperparametric Optimization. The result was observed by training two models on Cornell Movie-Dialogue Corpus, run by using 10 epochs. The comparison shows that after optimization, the model's accuracy and loss rate were 87% and 0.51%, respectively, compared to the results before optimizing the network (79% accuracy and 1.05% loss).

Keywords: Deep learning, Artificial Neural Network, Generative-based chatbot, hyperparameter optimization, Attentive Sequence-to-Sequence

1.0 INTRODUCTION

In 2016, 44% of consumers claimed they would rather interact with a chatbot than a human customer service agent [1]; and the percentages are projected to continue to rise in the future years. Every industry must design a solution that uses a third-party software to automate practically everything. Chatbot application naturally fit in with the characteristic of the industry demand. A chatbot is an application that is developed in the field of machine learning, which has become a hot topic of research in recent years. It's a Human-Computer Interaction (HCI) model and an artificial intelligence application that simulates a human-computer conversation [2]. Users' intent, input-output processing, and a response generating technique that obtains information from the knowledge base are all part of the application. Figure 1 depict the general architecture of the chatbot application from upper layer view.

The majority of today's chatbots integrate the Artificial Neural Network (ANN) approach with a Deep Learning environment. The combination of ANN and Deep Learning areas, as well as Natural Language Processing (NLP) techniques, results in a new generation chatbot known as a Generative-Based Chatbot. Generative-based methods leverage natural language generation (NLG) techniques to respond to a message [3][4]. However, to build a Deep Neural Network (DNN) Generative-Based chatbot requires a complex adaption of training network that works in tandem with optimal hyperparameter. As the environment can vary, how the hyperparameter reacts with the environment will also set a different optimum value. The current chatbot application mostly fails to recognize the optimum capacity of the network

Prediction in Smart Cities. *Smart Cities*. 3: 842–852. DOI: http://dx.doi.org/10.3390/smartcities3030043

[18] Jun, Q., D. Jun, S. Marco, M. Xiaoli, and L. Chin-Hui. 2020, August. On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression.

[19] Sebastian, R. 2018, September. *An overview of gradient descent optimization algorithms.*

[20] Mukkamala, M. C., and M. Hein. 2017. Variants of RMSProp and Adagrad with logarithmic regret bounds. *In Proceedings of the 34th International Conference on Machine Learning.* 70 (ICML'17): 2545–2553.

[21] Kingma, D. P., and J. Ba. 2019. Adam: A Method for Stochastic Optimization. *CoRR, vol. abs/1412.6980*.

[22] Senior, A., G. Heigold, M. Ranzato, and K. Yang. 2015. An empirical study of learning rates in deep neural networks for speech recognition. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 6724-6728. DOI: http://dx.doi.org/10.1109/ICASSP.2013.6638963

[23] Srivasta, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014, June. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*. 15: 1929 - 1958