2024, 21(2 Special Issue): 0655-0661
https://doi.org/10.21123/bsj.2024.9743
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

The Effect Of Optimizers On The Generalizability Additive Neural Attention For Customer Support Twitter Dataset In Chatbot Application

Sinarwati Mohamad Suhaili*[1,2], ***Naomie Salim***[1], ***Mohamad Nazim Jambli***[3]

[1]Faculty of Computing, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia.
[2]Centre of Pre-University, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.
[3]Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia.
*Corresponding Author.
PARS2023: Postgraduate Annual Research Seminars 2023.

Received 29/09/2023, Revised 10/02/2024, Accepted 12/02/2024, Published 25/02/2024

Abstract

When optimizing the performance of neural network-based chatbots, determining the optimizer is one of the most important aspects. Optimizers primarily control the adjustment of model parameters such as weight and bias to minimize a loss function during training. Adaptive optimizers such as ADAM have become a standard choice and are widely used for their invariant parameter updates' magnitudes concerning gradient scale variations, but often pose generalization problems. Alternatively, Stochastic Gradient Descent (SGD) with Momentum and the extension of ADAM, the ADAMW, offers several advantages. This study aims to compare and examine the effects of these optimizers on the chatbot CST dataset. The effectiveness of each optimizer is evaluated based on its sparse-categorical loss during training and BLEU in the inference phase, utilizing a neural generative attention-based additive scoring function. Despite memory constraints that limited ADAMW to ten epochs, this optimizer showed promising results compared to configurations using early stopping techniques. SGD provided higher BLEU scores for generalization but was very time-consuming. The results highlight the importance of finding a balance between optimization performance and computational efficiency, positioning ADAMW as a promising alternative when training efficiency and generalization are primary concerns.

Keywords: ADAM, ADAMW, Neural Network-based Chatbot, Optimizer, SGD.

Introduction

Integrating artificial intelligence (AI) through the use of neural networks is a widely used approach in various fields such as object and speech recognition, healthcare, and business, including chatbots. Chatbots based on neural networks typically aim to find the best function approximation by finding network parameters that minimize the error function during training data[1]. An error function measures how accurate the output of a model is compared to the actual output (target values). To improve the