

Project Proposal

ECE-9039-002

Team Members:

Yinglun Suo 251384723 ysuo7@uwo.ca

Zelin Zhang 251376484 zzha973@uwo.ca

Yanhua Zhang 251386851 yzha5778@uwo.ca

Jian Li 251391401 jli4627@uwo.ca

I. Background and Motivation

The advent of Natural Language Processing (NLP) has revolutionized how machines understand human language, with text classification being a pivotal task in this domain. Among the datasets facilitating this task, the AG News dataset stands out for its comprehensive coverage across four critical news categories: World, Sports, Business, and Technology.

Our project is motivated by the urgent need to address this imbalance. By aiming to develop an NLP model that not only excels in classifying text according to the AG News dataset's categories but does so with reduced computational resource consumption and fewer parameters, we tackle a critical issue. This endeavor is not just about achieving academic milestones; it's about pushing the boundaries of what's possible in deploying efficient NLP models in environments where computational resources are scarce or costly. Making high-performing, computationally efficient models could democratize access to advanced NLP technologies, especially in low-resource settings, thus broadening the impact of AI.

II. Research Objectives

Our primary objective is to engineer an NLP model that brings a new level of efficiency to text classification tasks, using the AG News dataset as a benchmark. This dataset, with its neatly categorized news articles, poses a unique challenge: how to maintain or improve classification performance, measured through AUROC and accuracy, while significantly reducing the model's demand on computational resources.

III. Methodology

Our research methodology is designed to systematically address the challenge of developing an efficient yet high-performing NLP model for text classification on the AG News dataset. The approach encompasses several key phases: dataset preparation, model design and optimization, training and validation, and performance evaluation.

- i. Dataset Preparation
- ii. Model Design
- iii. Training and Validation
- iv. Performance Evaluation

IV. Expected Outcomes

The anticipated outcomes of this research are twofold: a novel NLP model optimized for text classification on the AG News dataset with enhanced computational efficiency. This research seeks to push the boundaries of what is possible in NLP text classification by demonstrating that models can be both high-performing and computationally efficient. The success of this project could pave the way for the deployment of advanced NLP applications in resource-constrained environments, thereby democratizing access to cutting-edge AI technologies.

V. Timetable and Group Members

Task 1 [Data analysis, literature study and algorithm design] (week1)	
Determine Research Objectives	All members
Study the data sets and computing resources that we may obtain	
Learn some literatures from several top-level meetings of machine NLP	
Complete the proposal	Zelin Zhang
Task 2 [Network implementation] (week2 & week3)	
Design text preprocessing process	Jian Li
Design Neural Network	Zelin Zhang & Yinglun Suo
Set back propagation algorithm	
Refine and consolidate the components	Yanhua Zhang
Task 3 [Model training and fine tuning] (week4 & week5)	
Start training the model with a training set	Yinglun Suo
Adjust epoch and batchsize to maximize hardware resource utilization	Zelin Zhang
Use validation set to assess model accuracy	Jian Li
Adjust model hard variables and layers to improve accuracy and repeat the process	Yanhua Zhang
Task 4 [Report and Video] (week6)	
Completing Final Report	All members
Making presentation video	