PREDICTING THE CITATION COUNTS OF RESEARCH PAPERS USING NEURAL NETWORKS

I.M.D.Dayarathna

199314V

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

March 2021

DECLARATION

"I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books)."

UOM Verified Signature

Signature: Date: 29/05/2021

The supervisor/s should certify the thesis with the following declaration.

The above candidate has carried out research for the Masters thesis under my supervision.

Name of the supervisor: Dr. Charith Chitraranjan

UOM Verified Signature

Signature of the supervisor: Date: 29/05/2021

ABSTRACT

A widely accepted criterion used to measure the scientific impact of a research paper is the citation count. However, for a newly published paper this metric does not become available for several years after the date of publication. Yet, many parties including fellow scholars, research institutes and funding bodies find it important to be able to identify early on, the scientific papers with a higher potential to make a bigger impact. Predicting the future citation counts is an effective solution to overcome this limitation.

However, predicting the future citation count of a scientific paper is a challenging task, particularly due to the highly dynamic nature in the citation accumulation process. Hence this remains an active area of research. A majority of the prior studies that predict future citation counts using features available at the time of publication of a research paper make use of classical machine learning techniques. In this study, the author demonstrates through experiments how artificial neural network models can outperform best performing classical machine learning models discussed in prior studies.

One notable limitation of current approaches to this research problem is that many approaches treat the citation networks as unweighted graphs. In this work, the author demonstrates how treating the citation relationships as weighted relationships could help improve performance of the models. For this, the author introduces a novel feature named Weighted Average Neighboring Citation Score, a value computed by treating the citation network as a weighted graph, and demonstrates through multiple experiments that the newly introduced feature helps improve the performance of multiple models. Moreover, the author experiments with different edge weighting schemes and demonstrates how factoring both the recency of a citation and frequency with which a source has been cited when determining the edge weights help improve the performance of the models.

Keywords: Citation Count Prediction, Neural Networks, Weighted Citation Networks

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor, Dr. Charith Chitraranjan, for all the invaluable guidance provided to make this project a success.

I would also like to express my sincere gratitude to my family and friends for encouraging me on numerous occasions during this endeavor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

LIST OF ABBREVIATIONS

Abbreviation	Description
LR	Linear Regression
NN	Neural Network
SVR	Support Vector Regression
MLP	Multilayer Perceptron
GNN	Graph Neural Network

LIST OF APPENDICES