AUTOMATIC ANSWER GENERATION FOR MATH WORD PROBLEMS

Kulakshi Fernando

178090C

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science and Engineering

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

May 2019

DECLARATION

I declare that this is my own work and this dissertation 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 dissertation, 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).

Signature: Date:

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

Name of the Supervisor: Dr. Surangika Ranathunga
Signature of the Supervisor: Date:

Name of the Supervisor: Prof. Gihan Dias
Signature of the Supervisor: Date:

ACKNOWLEDGEMENTS

I am sincerely grateful for the advice and guidance of my supervisors, Dr. Surangika Ranathunga and Prof. Gihan Dias. Without their support, mentoring, and encouragement this project would not have been completed. I would like to thank them for taking time out of their busy schedule to be available anytime that was needed with help and advice.

I would also like to thank my progress review committee, Dr. Dulani Meedeniya and Dr. Supunmali Ahangama. Their valuable insights and guidance helped me to enhance the quality of my research work.

I would like to thank the entire staff of the Department of Computer Science and Engineering, both academic and non-academic for all their help during the course of this work and for providing me with any resources necessary to conduct my research.

This work was funded by a Senate Research Committee (SRC) Grant of the University of Moratuwa.

Finally, I would like to express my gratitude to my friends, specially Mr. Mokanarangan Thayaparan and Ms. Ishadi Jayasinghe who were always supporting me at my research work and to my family for their immense support.

ABSTRACT

Automatic Answer Generation for Math Word Problems

A math word problem (MWP) is a mathematical problem expressed using natural language. In this research, elementary level set-related word problems in which information is given in set notation are considered. As per our knowledge, this is the first research addressing set theory related word problems.

This research introduces an abstract representation to interpret mathematical semantics of set expressions and relations between sets. Two methods to extract given set related expressions were implemented: rule based method and a statistical method. Results show that statistical method is more robust to typing errors and unexpected expression formats. A parser based on a context free grammar is introduced to validate set related expressions and give feedback to the user when there are incorrect expressions. Along with these functionalities, we present a complete set problem solver system that understand and solve a given set word problem.

In addition to the solver, we experiment in extracting mathematical expressions from unstructured plain text using sequential classifiers. Several sequential classification models including conditional random-fields (CRF) and Long-Short Term Memory (LSTM) networks were compared with word and character level features. The results show that using character level features significantly increase the performance of mathematical expression extraction.

Keywords: Set theory. Answer generation. Math word problems.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

MWP	Math Word Problem
GCE	General Certificate of Education
IGCSE	International General Certificate of Secondary Education
GCSE	General Certificate of Secondary Education
JCE	Junior Certificate Examination
NCERT	National Council of Educational Research and Training
NLP	natural Language Processing
CRF	Conditional Random Field
LSTM	Long-Short Term Memory
Bi-LSTM	Bi-directional Long-Short Term Memory
CFG	Context Free Grammar

TABLE OF CONTENTS