AUTOMATIC MODEL ANSWER GENERATION FOR SIMPLE LINEAR ALGEBRA-BASED MATHEMATICS QUESTIONS

Rajpirathap Sakthithasan

168260N

Degree of Master of Science

Department of Computer Science & Engineering
University of Moratuwa
Sri Lanka

May 2018

AUTOMATIC MODEL ANSWER GENERATION FOR SIMPLE LINEAR ALGEBRA-BASED MATHEMATICS QUESTIONS

Rajpirathap Sakthithasan

168260N

Thesis 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 2018

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 acknowledgment 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.

.... ………………….
Sakthithasan Rajpirathap Date

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

....
Dr. Surangika Ranathunga Date

ABSTRACT

This research is focused on automating the process of generating answers to simple linear equation related mathematical problems.

Simple linear algebra based questions are a part of most Mathematics examinations. These linear algebra questions can appear as word type problems, where the question description is given in a textual form. Addition, subtraction, multiplication, division and ratio calculation are some of the known categories for linear equation based word type problems. Addition and subtraction based problems can be further divided based on their textual information as change type (join-separate type), compare type, and whole-part type. This research focuses on linear equation questions belonging to these three categories.

Mainly four approaches are followed by existing research for answer generation for linear algebra questions. These are rule/inference based, ontology based, statistical based, and hybrid based approaches.
In this research, a statistical approach is selected to automatically generate answers for simple linear algebra based model questions. The implemented system shows better accuracy than the other statistical systems reported in previous research for the same types of questions. This result is achieved by using ensemble classifiers and smart feature selection. Also, a new data set is created for training and evaluation purposes.

ACKNOWLEDGEMENTS

I would like to express profound gratitude to my adviser, Dr. Surangika Ranathunga, for her invaluable support by providing relevant knowledge, materials, advice, supervision and useful suggestions throughout this research work. Her expertise and continuous guidance enabled me to complete my work successfully.

I am grateful for the support and advice given by Dr. Malaka Walpola, by encouraging continuing this research till the end. Further, I would like to thank all my colleagues for their help in finding relevant research material, sharing knowledge and experience and for their encouragement. Also, I would like to thank for the teachers who involved in data creation, data annotation works in this research.

I am as ever, especially indebted to my parents for their love and support throughout my life. I also wish to thank my loving friends, who supported me throughout my work. Finally, I wish to express my gratitude to all my colleagues at my workplace, for the support given to me to manage my MSc research work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	**Description**
POS	Parts-of-speech
DOL	Dolphin Language
CFG	Context-free grammar
VBD	Verb, past tense
PRP	Personal pronoun
SVM	Support vector machine
NB	Naive Bayes
GCE O/L	General certificate of education ordinary level