USING MULTI AGENT TECHNOLOGY FOR AUTOMATIC MACHINE TRANSLATION

Budditha Hettige

(118036M)

Degree of Doctor of Philosophy

Department of Computational Mathematics

University of Moratuwa
Sri Lanka

July 2020

USING MULTI AGENT TECHNOLOGY FOR AUTOMATIC MACHINE TRANSLATION

Budditha Hettige

(118036M)

Thesis submitted in partial fulfillment of the requirements for the degree
Doctor of Philosophy

Department of Computational Mathematics

University of Moratuwa
Sri Lanka

July 2020

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

UOM Verified Signature

13.07.2020

………………………….   …………………

Signature: Date:

Budditha Hettige

Candidate

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

……………………………………………   …………………

Prof. Asoka S. Karunananda   Date:

UOM Verified Signature

14.07.2020

……………………………………………   …………………

Prof. George Rzevski   Date:

Dedicated to

This thesis is dedicated

...to my beloved mother and father

...to my wife and son

Acknowledgement

Many people have helped their best to successfully completion of this research. I acknowledge all of them for their valuable thoughts, and constant encouragement gave me to make my research a reality.

First and foremost, I acknowledge my supervisor senior professor Asoka Karunananda for accepting me as his research student and giving excellent support and advice. Prof. Karunananda is a great mentor who guided while giving me all the freedom and encouragement to accompany with my ideas. Also, I acknowledge my second supervisor, professor George Rzevski for accepting me as his research student and giving excellent support and advice. Without both them patient listening and creative thoughts, this work would not have been possible at all.

Especially, I acknowledge Venerable Kirioruwe Dhammananda Thero for his kind-hearted help and encouragement to fulfil my research work and give correct direction to my successful life.

My very special thank goes to Dr (Mrs.) Uditha Rathnayake, Dr (Mrs.) Menaka Ranasinghe and Dr Lochandraka Ranathunga for their invaluable comments and guidance as my examiners of bi-annual review panels.

Also, I wish to extend my sincere thanks for the support I received from all the members of the administration office and members of the Faculty of Information Technologies, University of Moratuwa. Especially I thank Dr (Mrs.) Thushari Silva and Dr (Mrs.) Subha Fernando and Madam, Prof. Deelika Dias for their essential roles.

I also acknowledge Mr P. Dias (Department of Statistics, University of Sri Jayewardenepura) and Ms. E.R.C. Sadamali for their kind support to fulfil my research work.

Also, exceptional and heartfelt thanks for Dr. (Mrs.) Mihirini Wagaarachchi, Dr (Mrs.) Chinthani Weerakoon, Dr. (Mrs.) Anuradha Ariyarthne and Ms. Mihiri Serisooriya for their gracious associations throughout the last couple of years.

I would like to thankfully remind all the academic and non-academic members of the General Sir John Kotelawala Defence University those who have supported me during the work. Special thanks go to the Vice-chancellor Major General Milinda Peiris, the Dean, Faculty of Computing, Commodore J. U. Gunaseela, for granting me the study leave to complete my research.

Again, I would like to mention Dr Asele Gunasekara, Maj R.M. Rathnayaka and all the members of my faculty gave me a great support

Finally, I would like to extend my greatest gratitude to my family members, especially my wife Lakshmi and my little hart Tenuja for the unrestricted support given, and without their care, this would have been unmanageable. Again, I must give an express thanks to Lakshmi for tolerating my busy schedules due to the research work. Last but not least, I thank all who supported me to make this work a success.

July 13, 2020 Budditha Hettige

Abstract

Machine translation is a cost-effective, quick, and widely accepted automated language translation method that has become essential in the modern and ever more globalized world. Machine translation can be done with one or more different approaches, including dictionary-based, rule-based, example-based, phrase-based, statistical, or neural-linguistic approaches. Nevertheless, most of the existing machine translation systems show a quality gap when compared with human translation. Thus, human translation has been considered as the best language translation method sofar. Human language translation is a complex and opportunistic process depends on human memory. This human language translation process has been described through a few theories. Among them, the garden path model and the constraint satisfaction model are two fundamental approaches available for human language translation, especially concerning sentence parsing with meaning. These two theoretical models demonstrate how to select suitable words in the phrase of a sentence to generate accepted meanings. Based on these two theories, a hybrid approach to machine translation has been proposed. This proposed approach is stimulated by how people parse and translate a sentence by putting available phrases together with accepted meaning. According to the approach, translation is done in three stages. In the first stage, the system analyses the given sentence by considering the morphology, syntax, and semantics of the source language. Then, the system uses phrase-based translation and translates each phrase into the target with multiple solutions. The phrase translation is done considering the four factors of psycholinguistic parsing techniques, such as phrase structure, semantic features, thematic roles, and probability. Finally, considering all the translated phrases, the system should be capable of identifying suitable target language phrases to take accepted meanings, considering subject-verb and object-verb agreements. After the subject-verb-object agreement, other available phrases in the sentence should be capable of re-arranging according to the accepted subject, object, and verb phrases.

This approach has been simulated with the multi-agent system named EnSiMaS, which translates English text into Sinhala. The EnSiMaS was implemented on the MaSMT framework, which was specially developed for agent-based machine translation. The EnSiMaS comprises of 26 language processing agents on both source and target languages. These agents were clustered into six agent swarms considering morphological, syntactical, and semantical concerns of the source and the target languages. In addition to these language-processing agents, the system should be able to create an agent dynamically for each source language phrase. These dynamically created phrase agents should be capable of communicating with other relevant phrases and taking the accepted solutions.

The EnSiMaS was tested with 85 sample English sentences. For each English sentence, three different translations were taken. According to the evaluation result, the system shows an 8.77% word error rate, a 6.72% inflexion error rate, and a 5.37% sentence error rate for the first, second, and third translations. In addition, calculated BLUE scores show 0.89160756, 0.52009204, and 0.43581893 for the first, second, and third translations. Then randomly selected 25 samples sentences are used to calculate the adequacy and fluency of the EnSiMaS. Adequacy and fluency rates were taken from 55 human evaluators considering the human-translated reference sentences. The Kendal's Tau correlation coefficient shows that there is a weak positive association between adequacy levels of human translations vs EnSiMaS system translations and moderate positive association between fluency levels of human translation and EnSiMaS system translation. Further, according to the Fleiss Kappa coefficient method, there is a significant fair agreement on raters for adequacy and fluency ratings.

Keywords: Machine Translation, Multi-agent systems, Human Language Processing, MaSMT, EnSiMaS

Table of Contents

List of Figures

.

List of Tables

List of Abbreviations

AI	- Artificial Intelligence
ARGM	- Agent Role Group Model
BCE	- Before the Current Era
BEES	- Bilingual Expert for English to Sinhala
CE	- Current Era
CAT	- Computer Assisted Translation
CYK	- Cocke–Younger–Kasami
CSM	- Constraint Satisfaction Model
EMA	- English Morphological Analysis
ESA	- English Syntax Analysis
ESMTS	- English to Sinhala Machine Translation System
EnSiMaS	- English to Sinhala Multi-Agent System
EBMT	- Example Based Machine Translation
FIPA	- Foundation for Intelligent Physical Agents
GNMT	- Google’s Neural Machine Translation
GPM	- Garden Path Model
HAMT	- Human-assisted (-aided) machine translation
IER	- Inflexion Error Rate
JADE	- Java Agent DEvelopment Framework
KQML	- Knowledge Query and Manipulation Language
LSTM	- Long Short Term Memory
LL	- Left-to-right, Leftmost derivation
LR	- Left-to-right, Rightmost derivation
MWE	- Multi-Word Expressions
MAS	- Multi-Agent System
MT	- Machine Translation
MaSMT	- Multiagent System for Machine Translation
NMT	- Neural Machine Translation
NLTK	- Natural Language Toolkit
NPMT	- Neural Phrase-based Machine Translation

NLP	- Natural Language Processing
PPO	- Preposition Phrase Order
RBMT	- Rule-based Machine Translation System
RTT	- Round-trip Translation
SL	- Source Language
SMT	- Statistical Machine Translation
SER	- Sentence Error Rate
SMG	- Sinhala Morphological Generation
SSG	- Sihala Syntax Generation
SOV	- Subject Object Verb
SVO	- Subject Verb Object
SPADE	- Smart Python multi-Agent Development Environment
TAG	- Tree Adjoining Grammar
TL	- Target Language
WER	- Word Error Rate