

# Automatic Fact Extraction from Open-Ended Geometry Questions

Ishadi Jayasinghe

188090K

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

November 2019

## DECLARATION

I, Ishadi Jayasinghe, 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 thesis/Dissertation under my supervision.

Name of Supervisor: Dr. Surangika Ranathunga

Signature of the Supervisor:

Date:

## ABSTRACT

**Semantic parsing of geometry problems** is the first step towards automated geometry problem solvers. Existing systems for this task heavily depend on language-specific NLP tools and use hard-coded parsing rules. Moreover, these systems produce a static set of facts and record low precision scores. In this study, we present the two-step memory network, a novel neural network architecture for deep semantic parsing of GWPs. Our model is language independent and optimized for low-resource domains. Without using any language-specific NLP tool, our system performs as good as existing systems. We also introduce on-demand fact extraction, where a solver can query the model about entities during the solving stage. This is impossible for existing systems; the set of extracted facts with these systems are static after the parsing stage. This feature alleviates the problem of having an imperfect recall.

We also investigate data augmentation techniques for low resource domains to alleviate the difficulties in applying deep learning techniques in the domain above. We also introduce an enhanced metric for evaluating language generative models alleviating the the limitations of exiting metrics. Analysing the results, we come up with a ranking of models on their suitability to be used o low resource domains

**Keywords:** Semantic Parsing; Deep Learning; Memory Networks; Generative Adversarial Networks; Temperature Sweep

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Surangika Ranathunga for the continuous support given for the success of this research, for her patience, motivation, and immense knowledge. I'm extremely thankful for your advices. I strongly believe that without your tremendous mentorship and advice from the beginning, the present work could have not reached this stage. Your wide knowledge and logical way of thinking have been of great source of inspiration for me.

I wish to thank Ms. Vishaka Nanayakkara and Mr. Nisansa de Silva for their valuable insights and guidance from the very early stage of this research. 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 the resources necessary to conduct my research. This research was supported by the University of Moratuwa Senate Research Grant.

I would like to thank Mr. Moksha Manukantha, Ms. Kulakshi Fernando, Ms. Vijini Liyanage, and my family for all the love and support.

**Thank you!**

## LIST OF ABBREVIATIONS

NLP	Natural Language Processing
GWP	Geometry Word Problem
MCQ	Multiple Choice Question
NLG	Natural Language Generative
DNN	Deep Neural Network
GAN	Generative Adversarial Net
CNN	Convolutional Neural Network
RNN	Recurrent Neural Networks
LSTM	Long Short Term Memory
GRU	Gated Recurrent Unit
DBN	Deep Belief Net
DAE	Denoising Autoencoder
VAE	Variational Autoencoder
MLE	Maximum Likelihood Estimate
RL	Reinforcement Learning
FC	Focus Controlling
PE	Position Encoding

## LIST OF FIGURES

Figure 1.1	An example for a geometry proof problem	2
Figure 1.2	An example for an MCQ in geometry	2
Figure 1.3	Dependency parse tree for “AB is a chord of the circle with center O” using Stanford CoreNLP Toolkit [1]	4
Figure 2.1	Attention mechanism introduced by Bahdanau et al. [2]. Image source: [2]	11
Figure 3.1	(a) Two examples NLG models named A and B plotted on a graph with the x-axis as self-BLEU and the y-axis as test-BLEU. (b) Two example temperature curves of two NLG models. The axes have the inverse metrics, so the lower is better in both axes. When it comes to the model with the blue curve, in any given quality, it has better diversity than the model with the red curve. Therefore, this graph is sufficient enough to claim that the model with the blue curve is better.	22
Figure 4.1	Architecture of the sentence selection model. The lighter the color, the higher the probability. This design relates to the single hop scenario, which is sufficient for the task we focus on. However, for tasks needing to handle inter-dependencies between sentences (such as coreference resolution), we can adopt the same word-level model (figure 4.2) with having sentences instead of words.	25
Figure 4.2	Architecture of the Word-level memory network	26
Figure 4.3	Querying the sentence “In the figure above, line AB, line CD, and line EF intersect at P.” with the keyword “line”.	31
Figure 4.4	Binary rule extraction for the keyword “lies” from the sentence “In the figure above , point O lies on line AB.”. Here we can see how the the two literals ( “O” and “AB”) are with high probabilities in the last two layers.	32

Figure 4.5	Fact extraction for the entity “m” from “In the figure above, lines l and m are not parallel.” Lighter colors indicate higher probabilities	33
Figure 4.6	Fact extraction for the entity “v” from “In rectangle ABCD above, the area of the shaded region is given by v”. Here, we can see how the answer “area” is refined over hops. If there were only a single hop, the answer would have been the word “by” (word with the lightest color in row 1) which is incorrect.	34
Figure 5.1	Area under the temperature curve of COCO dataset plotted with respect to (a) selfBLEU axis (b) (1− testBLEU axis) axis. Area under the temperature curve of EMNLP NEWS dataset plotted with respect to (c) selfBLEU axis , and (d) (1− testBLEU axis). The area w.r.t. selfBLEU axis is inversely proportional to the quality of the generated sample. Thus, higher values reflect poor quality samples, and vice versa. Similarly, the area w.r.t.(1− testBLEU) axis is inversely proportional to the diversity of the sample. Thus, higher values reflect samples with lesser diversities.	44
Figure 5.2	Experiment results on MS COCO dataset	45
Figure 5.3	Experiment results on EMNLP NEWS dataset	46

## LIST OF TABLES

Table 4.1	Task Definition	24
Table 4.2	Statistics of the dataset. Introduced by Seo et al. [3]	30
Table 4.3	Precision, Recall, and F1 score for the task of unary rule extraction	30
Table 4.4	Precision, Recall, and F1 score for the task of binary rule extraction	32
Table 4.5	Precision, Recall, and F1 score for the task of on-demand fact extraction	34
Table 4.6	Precision, Recall, and F1 score for relation completion	35
Table 5.1	Summary statistics of the datasets used	43



# TABLE OF CONTENTS

Declaration of the Candidate & Supervisor	i
Abstract	ii
Acknowledgement	iii
List of Abbreviations	iv
List of Figures	v
List of Tables	vii
Table of Contents	viii
1 Introduction	1
1.1 Background	1
1.2 Research Problem	2
1.3 Research Objectives	5
1.4 Contributions	5
1.5 Publications	6
2 Background	7
2.1 Overview	7
2.2 Recurrent neural networks	7
2.3 Attention Mechanisms	10
2.4 Memory Networks	11
2.5 End-to-end Memory Networks	12
2.6 Generative Adversarial Nets	14
3 Literature Survey	16
3.1 Automated solving and assessment of geometry problems	16
3.2 Relation Extraction	17
3.3 Text generation	18
3.4 Evaluation of text generation models	20
4 Deep semantic parsing of geometry problems	23
4.1 Task Definition	23
4.2 Limitations of existing memory networks	23

4.3	Model Formulation	24
4.3.1	Focus controlling (FC) with position encoding	27
4.3.2	Unary Rule Extraction	28
4.3.3	Binary Rule Extraction	28
4.3.4	On-demand Fact Extraction	28
4.4	Experiments	29
4.4.1	Unary Rule Extraction	29
4.4.2	Binary Rule Extraction	32
4.4.3	On-demand Fact Extraction	32
4.4.4	Relation Completion	34
4.5	Discussion	35
5	Text Generation on low-resource domains	37
5.1	Task Definition	38
5.2	Experiments	38
5.2.1	Enhanced metric for evaluating NLG models	38
5.2.2	Experiment Setup	39
5.2.3	Experiments with MS COCO captions	41
5.2.4	Experiments with EMNLP2017WMT News Dataset	45
5.2.5	Experiments with Geometry Questions	46
5.3	Discussion	47
6	Conclusion and Future work	49
	References	51