**Research Proposal**

Extension of Neural Reasoner for Question Answering

Student ID: 26250365

Name: Chenyang Wang

Course: Bachelor of Information Technology (Honours)

Date: 7/04/2017

Supervisor: Reza Haffari

Contents

[Introduction 3](#_Toc479334158)

[Background 4](#_Toc479334159)

[Research Questions 7](#_Toc479334160)

[Main Goal 7](#_Toc479334161)

[Sub-goals 7](#_Toc479334162)

[Research Design 8](#_Toc479334163)

[Methodologies 8](#_Toc479334164)

[Proposed Chapters 10](#_Toc479334165)

[Timetable 11](#_Toc479334166)

[Potential Difficulties 12](#_Toc479334167)

[Special Facilities Required 12](#_Toc479334168)

[Expected Outcomes 13](#_Toc479334169)

[References 14](#_Toc479334170)

# Introduction

In recent years, there has been great advancements in applying neural networks to natural language processing tasks, including word representation [1] [2], dependency parsing [3] [4], sentiment analysis [5], etc. A new use case that has seen much development lately is question answering. The basic form of the task is as follows: given a corpus of text and a question sentence, extract relevant information from the corpus and find a sensible answer word or sentence to the question. In a traditional NLP workflow, this is usually done through the joint use of a rule-based system, which reduces sentences to basic logic forms and performs inferences on them, and some search-based techniques. This has the drawback of having to manually design rules and features for the system, and is not flexible enough compared to the complexity of natural languages.

Neural network-based solutions have the advantage of being end-to-end and capable of learning syntactical as well as semantical relationships in natural languages. Recently there has been several notable efforts on question answering, including using recursive neural networks [4] and neural tensor networks [6] for sentence summarisation, using dynamic memory networks (DMN) with attention mechanisms to extract answers [7], and using multiple layers of neural reasoner units to find an answer from multiple facts [8]. In this research project, we will be examining the neural reasoner approach while incorporating some of the features of the DMN model, focusing first on explaining the properties of the neural reasoner unit, and then based on what we learned, on improving the performance of the system by incorporating more expressive sentence representation techniques and attention mechanisms into the neural reasoner framework. The expected contributions of this paper will be:

* An intuitive understanding as well as a formulated description of what each neural reasoner unit is performing in the reasoning iterations, allowing further work to be done on:

1. Optimising word and phrase representations
2. Adjusting training objectives
3. Individually (pre-)training neural reasoner units, as well as using a neural reasoner unit separately for information extraction
4. Evaluating number of iterations required for the reasoning process

* An improved neural reasoner framework incorporating recent discoveries from related efforts that outperforms the original neural reasoner network and generalise better on more diverse tasks.

# Background

Two of the most relevant recent research to our work are the neural reasoner model [8] and the dynamic memory network [7].

Our work will be mainly based on [8] in which a new neural reasoner framework is proposed. In the input layer, the network passes word embeddings of question and fact sentences through a bidirectional recurrent network to obtain fixed-length phrase representations. Then in the reasoning layers, it iteratively pairs the question representation with each of the fact representations and pass them through a deep network unit, outputting a revised question vector and a revised fact vector. All new question vectors are then combined through a pooling layer to form the question vector of the next iteration. The answer is then extracted from the final question vector by an answerer network after several iterations. This approach resembles the process of a traditional resolution-based logical inference.

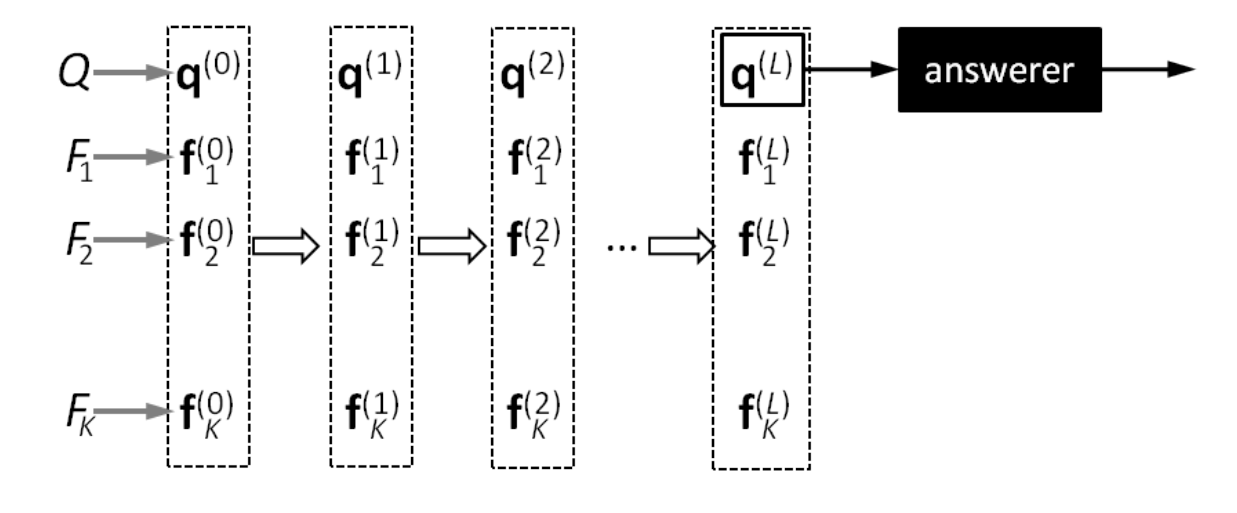


Figure 1. Structure of the Neural Reasoner (Peng, et al. 2015)



Figure 2. Basic Unit of the Reasoning Layers

A similar model is outlined in [7] where a dynamic memory network (DMN) is used. The Idea of memory networks is first proposed in [9] for question answering tasks, in which Kumar et al. expands the idea of memory networks to allow the model to work with sequential input. Unlike the neural reasoner model, the input sentences are treated as an entire input sequence (thus taking into account the order of sentences), and is passed through a recurrent neural network. The hidden states are used as inputs to the episodic memory module. In multiple iterations, an attention mechanism selects the most relevant fact representation, combine it with the question representation and a previous memory to update an individual memory (called an episode) for that input, and the episode is used to update the overall memory for that iteration. The answer is extracted from the final episodic memory by the answer module. The main difference between the DMN and neural reasoner is that:

1. The input facts are treated as an entire sequence.
2. The fact and question representations are not transformed after each iteration. Instead, they are used to update a corresponding *episode* for each fact.

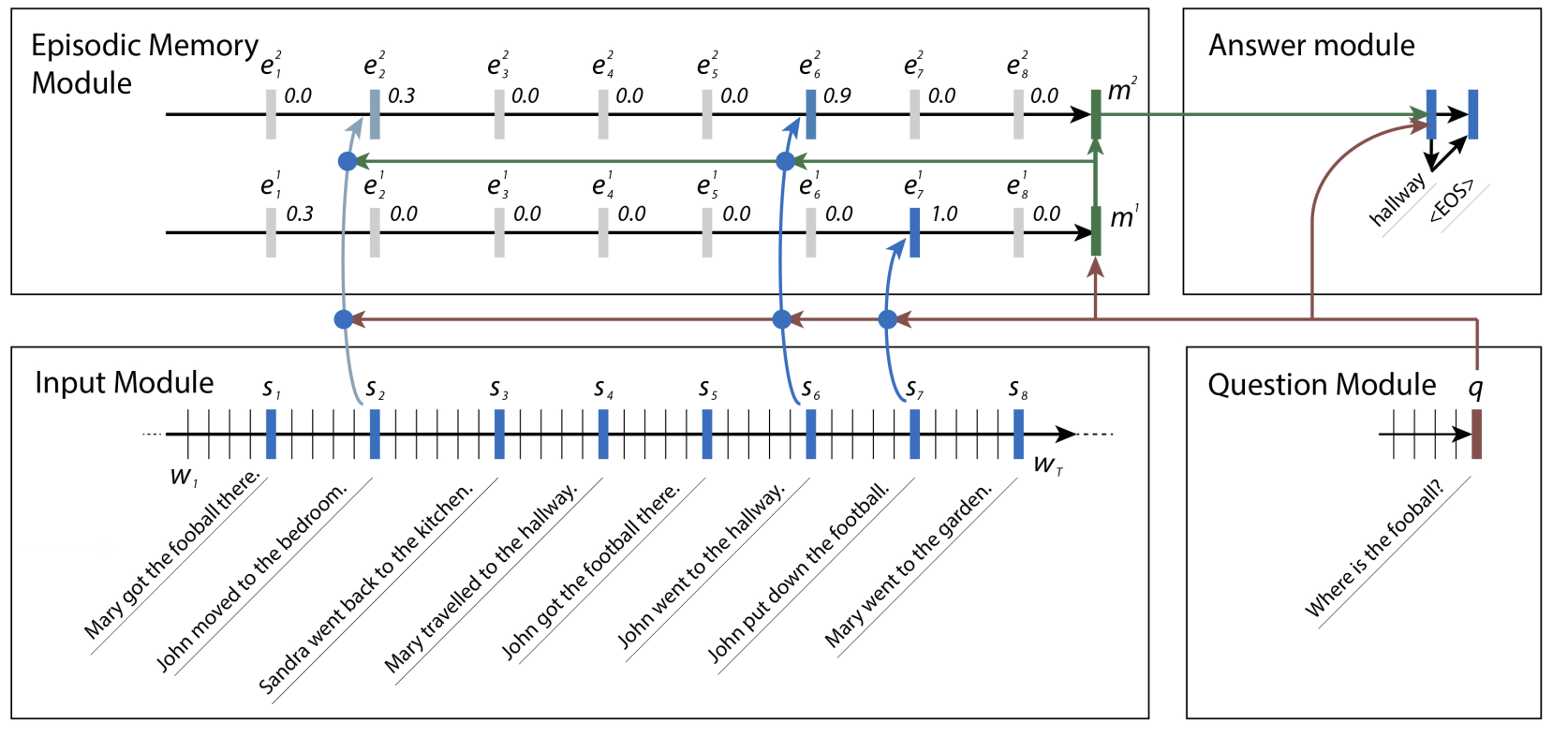


Figure 3. Structure of the DMN model (Kumar, et al., 2015)

It is reported in [8] that the neural reasoner outperforms the memory network and DMN model significantly in larger datasets for two position-related tasks in the bAbI task set. However, in [7] it is reported that the DMN model performs well on all other 18 tasks out of the 20 tasks in the bAbI dataset, which are not yet well-tested on the neural reasoner model. We suspect that in order to achieve high performance on as many types of tasks as possible, a combination of features from both the neural reasoner and DMN model are needed.

There have been recent efforts to extend the neural reasoner or DMN model. In [10] Yin et al. uses multiple layers of convolutional networks to generate representations of not only sentences but also snippets of sentences, and then use an attention mechanism to generate a query-biased representation of the input for answer extraction. This model obtains better performance on the MCTest dataset [11] than several baseline models including the neural reasoner. However, it is unclear whether their attention-based convolutional network performs well in datasets where DMN and neural reasoner demonstrate high performances. Further investigation is needed to assess the benefits of adopting a multi-sentence level representation.

Several other related research works are of interest for improving the design of network components in existing neural reasoning models. In [6] a neural tensor network (NTN) is used to model reasoning where the relationship between two entities is involved. It is shown that a simple bilinear composition of two input vectors can be used to model multiple types of interactions between two entities while remaining relatively easy to train for the network. Despite its original use in modelling entity relationships, the NTN can also potentially be used to model phrases from word representations, as well as to model the interaction between questions and facts. However, it is shown in [3] that in given sentence modelling tasks, a dynamic convolutional network (DCNN) outperforms the NTN. In addition, the DCNN has the advantage of potentially not requiring the input sentences to be pre-parsed. This suggests a DCNN can possibly replace the RNN or bidirectional RNN in the DMN or neural reasoner model as a better way to model input questions and facts.

Also crucial to this project is the understanding of word embeddings and the composition/reasoning networks as operators in the embedding space. [1] and [2] introduce general ways of creating word embedding vectors with both syntactical and semantical significance. Along with the technique of creating dense representations of words is the idea to use distance of representations in the embedding space to measure their similarities and to use linear operations to calculate word analogies. Techniques such as PCA, RCA or t-SNE [12] can be used to illustrate relationships in the embedding space and has already been used to demonstrate the effectiveness of using dense vector representations for word analogy tasks [13]. The same technique can be adopted to study the properties of neural networks that transform one set of representations into another, such as the reasoning layer of the neural reasoner.

In [14] Chen and Manning proposed a transition-based algorithm for dependency parsing of sentences, effectively transforming the task from a sequence-to-tree mapping to a reinforcement learning problem. While not directly related to question answering, the concept of transition-based algorithms can be applied to question answering as well, considering the iterative reasoning step in DMN and neural reasoner models involve well-defined steps such as selecting facts, combining facts with question and terminate reasoning. A reinforcement learning-based algorithm can potentially be used to learn to exit the reasoning iterations on its own. This is a promising direction for further investigation.

A benchmark for question-answering models, the bAbI dataset, is established in [15]. 20 tasks with different types of required reasoning are given, and several baseline models as well as state-of-the-art models at that time are tested. This offers us a high-quality tool for evaluating the performance of our models versus existing solutions.

# Research Questions

## Main Goal

The main goal of this project is to understand the properties of the neural reasoner and improve the performance of the neural reasoner network on selected benchmarks.

## Sub-goals

1. Understand the reasoning process and obtain an intuitive as well as formal explanation of the process. Test the hypothesis that each neural network module in the reasoning layers performs the task of relevant fact separation.
2. Look for a more expressive and information-preserving method for generating phrase and sentence embeddings.
3. Incorporate an attention mechanism into the reasoning layers to improve their effectiveness.
4. Discuss possible conditions for terminating the reasoning iterations.

# Research Design

## Methodologies

The original neural reasoner network outlined in [8] will be reimplemented in TensorFlow as a baseline. All following variants and improvements of the model will also be implemented in TensorFlow version 1.0 in Python 3. The initial benchmarking dataset will be the CNN and Daily Mail Question Answering Dataset proposed in [16]. In [17] it is suggested that current reading comprehension algorithms may have approached their maximum performance on this dataset, therefore for further improvement of the algorithm, other benchmarking datasets will be required. The bAbI dataset [15] used in [8] originally will be used in conjunction of the CNN/DM dataset for benchmarking.

For understanding the properties of the neural reasoner, the original model will be trained on the bAbI dataset until reaching similar performance level in the original paper. Several test cases will then be passed through the network and the intermediate question and fact representations will be recorded. The relative location of the original facts, question, intermediate representations, and target answer in the embedding space will be visualised with PCA, RCA or t-SNE (described in [12]). The same procedure will be repeated on an individual reasoning network module to observe the relationship between the inputs and outputs. Visual evidence will be used to test the hypothesis that each individual reasoning network learns to separate relevant facts and deductions from irrelevant facts. Potential findings in the visualisation will be validated by distance statistics on a larger sample of the dataset.

For improving phrase representations, the recursive neural network outlined in [4] will be used to generate question and fact representations. For simplicity, a recursive network on pre-parsed sentences will be implemented first to evaluate its relative performance versus the bidirectional recursive network encoder in [8]. If the results are positive, a full structure-learning recursive network algorithm will be implemented for phrase representation encoding.

For incorporating an attention mechanism into the reasoning layers, we will be studying the attention mechanism used in [7]. As a first proposal, the attention-based update rule will be used to update intermediate question and fact representations in the reasoning layers and it will be benchmarked against the original model.

Evidences from visualisation and distance statistics analysis of the phrase embedding space in the earlier step will be used in search of a plausible terminating condition of the iterative reasoning process.

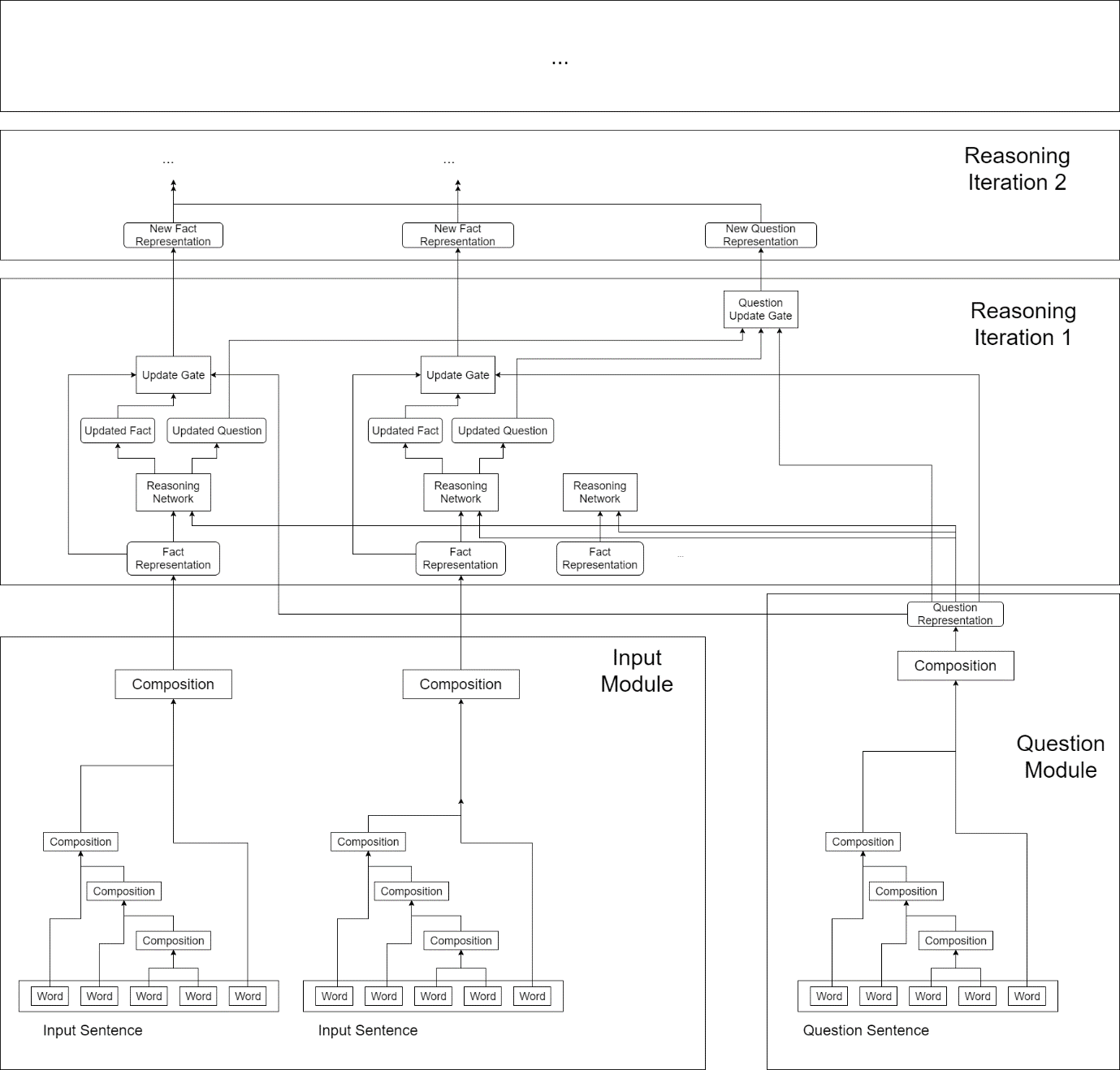


Figure 4. One Possible Configuration of the Improved Neural Reasoner Network

## Proposed Chapters

|  |  |  |
| --- | --- | --- |
| Chapter Name | Est. Word Count | Purpose |
| Abstract | 150 | Brief outline of the research project. |
| Introduction | 500 | Explain the background, purpose and aim of the research, as well as briefly demonstrate the main results. |
| Related Work | 1000 | Discuss relevant prior works and notable ideas and findings closely related to the research. |
| Neural Reasoner | 1500 | Explain the idea of neural reasoner. Discuss the intuition and justifications of the design. Discuss potential points of improvement. |
| Reasoning Process Visualisation | 2500 | Propose a way of embedding space visualisation. Visualise the effect of neural reasoner on question and fact embeddings. |
| Word and Phrase Representations | 3000 | Explain different word and phrase embedding techniques and their strengths/weaknesses. Integrate different word embedding techniques into the original neural reasoner. |
| Attention Mechanism | 3000 | Briefly explain the use of attention mechanism in NLP. Integrate an attention mechanism into the neural reasoner model. |
| Experiments | 1000 | Outline the results from representation and attention mechanism experiments. |
| Discussions | 2000 | Explain the outcomes of the experiments. Reach a conclusion. |
| Further Research | 500 | Propose potential further research ideas. |
| References | - | - |
| Appendix | - | Neural network implementation details, additional graphs, etc. |
| (total words) | ± 15000 |  |

## Timetable

|  |  |  |
| --- | --- | --- |
| Time | Research Work | Thesis Progress |
| Semester One |  |  |
| Week 6 onwards | Further literature review | Research proposal |
| Week 8 | Reproducing the original neural reasoner | Start writing literature review |
| Week 9 | Train the model on the benchmark dataset |  |
| Week 10 | Attempt visualisation of the embedded space | Document implementation process |
| Week 12 | Visualise the effects of the neural reasoner | Organise visualisation results in a report |
| Week 13 |  | Prepare for presentation |
| Week 14 | Conclude neural reasoner property analysis | Literature review & interim presentation |
| Semester Break |  |  |
| First Half | Implement alternative phrase embeddings without structure learning |  |
| Semester Two |  |  |
| Week 2 | Implement alternative phrase representation with structure learning |  |
| Week 4 | Phrase representation performance test | Organise performance test results in a report |
| Week 5 | Find and train the best model so far |  |
| Week 7 | Attempt layer by layer training | Document implementation process |
| Week 10 | Experiment with attention mechanism | Document implementation process |
| Week 12 | Draw insights and look for convergence conditions (stretch goal) | Thesis review |
| Week 13 |  | Prepare for presentation |
| Week 14 |  | Final presentation |
| Week 15 | Finalise thesis | Final thesis |

## Potential Difficulties

1. There can be difficulties implementing the model in the original neural reasoner paper due to lack of clarification on implementation details such as hyperparameters or optimiser settings.

**Mitigation:** Allocate more time for model training and tuning. Incorporate latest neural network training techniques into model training. Train multiple models with different settings in parallel

1. Finding an appropriate data set for training and benchmarking models is important. It is necessary to evaluate the effectiveness of a data set before using it for the research.

**Mitigation:** Read reviews of existing benchmarking datasets. Evaluate the effectiveness of several benchmarks and select 2-3 benchmarks most relevant to this research.

1. It is still unclear whether visualisation of the embedded space can provide meaningful insights into the inner workings of the neural reasoner. If it is unable to provide us with a better interpretation of the neural reasoner, alternative techniques will be required to discover its properties.

**Mitigation:** Fall back to simpler statistical measures or investigate the convergence of outputs after iteration for analysis

1. Training and evaluating the models require a large amount of computing power. At later stages of this research, specialised computing nodes or similar cloud solutions may be needed.

**Mitigation:** Request for cloud-based compute engine access. Utilise Google Cloud free trial.

## Special Facilities Required

At the implementation and performance testing stages of the project, A sufficiently fast compute node or a similarly powerful cloud-based virtual machine is required to complete the training and testing of different models as well as performing inferences on multiple evaluation tasks.

# Expected Outcomes

## Deliverables

1. An informative explanatory visualisation of the word embedding space illustrating of the neutral reasoner's operations on input question and fact representations, as well as a discussion on its properties and strategies to exploit them.
2. An improved network model with better input encoding techniques that take into account the structure of input phrase or sentences, and an attention mechanism to improve the efficiency of reasoning iterations.
3. A benchmark report comparing the original neural reasoner model and our modified variants on 2-3 different tasks. It should be able to demonstrate the advantages of adopting our modifications.
4. An algorithm or strategy to determine the number of necessary iterations to reach a confident answer given a specific task.

## Contributions

The visualisation techniques in this project will provide a novel way to understand the actual calculations performed by the network to solve reasoning tasks, as well as a method to diagnose potential issues in the network. Such techniques may also assist the design of new network configurations that perform different types of operations in the word embedding space by allowing the designer to directly observe the effects.

An extended neural reasoner model from this research will further improve the performance of neural network-based models in question-answering tasks, especially on tasks with non-trivial sentence and inter-sentence structures. This is important, as most real-world question-answering problems, unlike synthetic benchmarks, have more complicated and irregular word and sentence arrangements. Being able to cope with these irregularities by exploiting the structure of the language will significantly open up more applicable use cases of a question-answering model.

# References

|  |  |
| --- | --- |
| [1] | T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean., “Distributed representations of words and phrases and their compositionality.,” *Advances in neural information processing systems,* pp. 3111-3119, 2013. |
| [2] | J. Pennington, R. Socher and C. D. Manning, “Glove: Global Vectors for Word Representation.,” *EMNLP,* vol. 14, pp. 1532-1543, 2014. |
| [3] | N. Kalchbrenner, E. Grefenstette and P. Blunsom, “A Convolutional Neural Network for Modelling Sentences,” *arXiv:1404.2188,* 2014. |
| [4] | R. Socher, C. C. Lin, C. Manning and A. Y. Ng, “Parsing natural scenes and natural language with recursive neural networks.,” in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011. |
| [5] | R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng and C. Potts, “Recursive deep models for semantic compositionality over a sentiment treebank.,” *Proceedings of the conference on empirical methods in natural language processing (EMNLP),* vol. 1631, p. 1642, 2013. |
| [6] | R. Socher, D. Chen, C. D. Manning and A. Ng., “Reasoning with neural tensor networks for knowledge base completion.,” *Advances in neural information processing systems,* pp. 926-934, 2013. |
| [7] | A. Kumar, O. Irsoy, J. Su, J. Bradbury, R. English, B. Pierce, P. Ondruska, I. Gulrajani and R. Socher, “Ask me anything: Dynamic memory networks for natural language processing.,” *CoRR, abs/1506.07285,* 2015. |
| [8] | B. Peng, Z. Lu, H. Li and K.-F. Wong, “Towards neural network-based reasoning.,” *arXiv:1508.05508,* 2015. |
| [9] | J. Weston, S. Chopra and A. Bordes, “Memory networks.,” *arXiv:1410.3916,* 2014. |
| [10] | W. Yin, S. Ebert and H. Schütze, “Attention-based convolutional neural network for machine comprehension.,” *arXiv:1602.04341,* 2016. |
| [11] | “https://www.microsoft.com/en-us/research/publication/mctest-challenge-dataset-open-domain-machine-comprehension-text/,” Microsoft, 1 October 2013. [Online]. |
| [12] | L. v. d. Maaten and G. Hinton, “Visualizing data using t-SNE.,” *Journal of Machine Learning Research,* pp. 2579-2605, 9 2008. |
| [13] | J. Pennington, R. Socher and C. D. Manning, “GloVe: Global Vectors for Word Representation,” 2014. [Online]. Available: https://nlp.stanford.edu/projects/glove/. [Accessed 5 April 2017]. |
| [14] | D. Chen and C. D. Manning., “A Fast and Accurate Dependency Parser using Neural Networks.,” *EMNLP,* pp. 740-750, 2014. |
| [15] | J. Weston, A. Bordes, S. Chopra, A. M. Rush, B. v. Merriënboer, A. Joulin and T. Mikolov, “Towards ai-complete question answering: A set of prerequisite toy tasks.,” *arXiv:1502.05698,* 2015. |
| [16] | K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman and P. Blunsom, “Teaching machines to read and comprehend.,” *Advances in Neural Information Processing Systems,* pp. 1693-1701, 2015. |
| [17] | D. Chen, J. Bolton and C. D. Manning, “A thorough examination of the cnn/daily mail reading comprehension task.,” *arXiv:1606.02858,* 2016. |