IMPROVEMENT OF WORD EMBEDDINGS

NEURAL REASONING MODEL FOR QUESTION ANSWERING

18-090

Preliminary Progress Review

(Preliminary Progress Review Documentation submitted in partial fulfilment of the requirement for the Degree of Bachelor of Science Special (Honours) In Information Technology)

A.K.R.R. Aneeze (IT15060372)

Bachelor of Science (Honours) in Information Technology
(Specialization in Software Engineering)

Department of Information Technology

Sri Lanka Institute of Information Technology

May 2018

1.0. INTRODUCTION

Question Answering Systems are a popular application of automated linguistic analysis techniques. These systems use advanced language modeling and feature learning techniques to generate state of the art results. These techniques are collectively referred to as word embeddings.. Being the initiating point for a lingual system by modeling the inputs for the neural network, word embedding contributes to the precision and accuracy of the whole system. Methodologies to generate these word embeddings are diverse probabilistic models, dimensionality reduction and neural network mappings being few examples. Out of these neural network based models have shown promising results over the years due to their ability to learn subtle linguistic features and use the knowledge to evolve over time.

This document looks into the techniques used for word embedding, the proposed methodology to improve word embeddings and the anticipated benefits.

1.1. Purpose

The purpose of this document is to provide the preliminary progress of the "Improvement of Word Embedding" component of "Neural Reasoning Model for Question Answering".

This document is primarily intended to be proposed to the audience who has already referred to the proposal of "Neural Reasoning Model for Question Answering" and also to anyone who has a knowledge on word vector modeling. Further, this document will be used as a reference for the progress of the component.

1.2. Scope

The document will illustrate the purpose and the main areas to be focused throughout the component. The literature review section will demonstrate the research gap that this project is attempting to fulfill by analyzing the existing solutions in the QA domain, revealing the advantages and disadvantages of those systems, and how the proposed system can improve on those failings. Furthermore, the research objectives and the proposed methodology will be discussed. Latter part of the document will contain the data sources, anticipated benefits and the anticipated deliverables for this project.

1.3. Overview

The proposed system aims to achieve a remarkable improvement in the Reasoning, Attention and Memory(RAM) Model in Deep Neural Networks, thus achieving effective and accurate answers in the Question Answering system. The main goal of the reasoning component would be to allow the system to reason about entities and relationships similar to a human.

2.0. STATEMENT OF WORK

2.1. Background Information and Overview of Previous Work based on Literature Survey

Automated QA systems have been gaining a lot of prominence since the early 60's. QA has mostly been used to develop intricate dialogue systems such as chat-bots and other systems that mimic human interaction [2]. The methodologies used in these systems vary from Information retrieval based statistical approaches to machine learning approaches and then to deep learning based evolving systems.

Traditionally, most of these systems use the tried methods of parsing, part-of-speech tagging, etc that come from the domain of NLP research. While there is absolutely nothing wrong with these techniques, they do have their limitations. [3] W.A. Woods et al. shows how we can use NLP as a front end for extracting information from a given query and then translate that into a logical query which can then then be converted into a database query language that can be passed into the underlying database management system. In addition to that there needs to be a lexicon that functions as an admissible vocabulary of the knowledge base so that it is possible to filter out unnecessary terminology. The knowledge base is processed to an ontology that breaks it down into classes, relations and functions [4]. Natural Language Database Interfaces (NLDBIS) are database systems that allow users to access stored data using natural language requests. Some popular commercial systems are IBM's Language Access and Q&A from Symantec [5].

Information retrieval (IR) is another technique that has been used to address the problem of QA. With IR systems pay attention to the organization, representation and storage of information artifacts such that when a user makes a query the system is able to to return a document or a collection of artifacts that relate to the query [6]. Recent advances in OCR and other text scanning techniques have meant that it is possible to retrieve passages of text rather than entire documents. However IR is still widely seen as from the document retrieval domain rather than from the QA domain.

Template based question answering is another technique that has been used for QA and is currently being used by the START system which has answered over a million questions since

1993 [7]. START uses natural language annotations to match questions to candidate answers. An annotation will have the structure of 'subject-relationship-object' and when a user asks a question, the question will be matched to all the available annotation entries at the word level (using synonyms, IS-A, etc) and the structure level. When a successful match is found, the annotation will point to an information segment which will be returned as the answer. When new information resources are incorporated into the SMART system, the natural language annotations have to be composed manually [8]. START uses Omnibase as the underlying database system to store information and when the annotation match is found, the database query must be used to retrieve the information. While this system has been relatively successful, it requires a lot of preprocessing which must be done manually.

Deep learning is the state-of-the-art in areas such as speech recognition, natural language understanding, visual object recognition, etc. Convolutional neural networks have brought many breakthroughs in areas such as processing images, pictures and speech, whereas Recurrent networks have been extremely successful in areas such as processing text and speech.

Recurrent Neural Networks are special because of its ability to take the previously perceived information into consideration. While its counterpart Feed-forward neural networks are only concerned with the state of the inputs at a given time, RNN takes the past decisions into consideration. The decision a recurrent net reached at time step t-1 affects the decision it will reach one moment later at time step t. This is because of the feedback loop that enables RNNs to ingest their own outputs moment after moment as input retaining memory in the sequence itself.

The sequential information in an RNN can be preserved in the hidden state of the network and it can be carried forward several steps so that it would affect processing new inputs. This enables RNN to find co-relations between inputs that are separated by many moments. As long as the memory of a recurrent net can persist the past decisions of would influence the current decision making.

One of the major fallbacks of the traditional RNNs is the vanishing gradient problem. These RNNs were not able to form connections between the final outputs and the events that happened several steps before. Because of this the gradient for that particular period cannot be calculated which resulted in the “vanishing gradient” for that period in the graph. One reason for the vanishing gradient is that the information passing through the network passes through several stages of multiplication. If we were to explain this problem using basic mathematics, if a value is to be multiplied by another value more than 1(even slightly) continuously it can become immeasurably large. In the same way if the multiple is less than 1, the value tends to do the inverse. The first example causes Exploding Gradients but this can be solved easily by truncating. But the vanishing gradient caused by the second example is harder to solve because it can become too small for computers to work with.

To solve this problem a variation of RNN came up with Long Short-Term Memory units, or LSTMs. LSTMs preserve a constant error that can be that can be back-propagated through many time and layers, sometimes as many as 1000. This opens up channels to link cause and effect remotely. LSTM achieves this by storing the information outside the normal flow in a gated cell. The cell makes decisions about what to store, and when to allow reads, writes and erasures, via gates that open and close. These are analog gates implemented with element-wise multiplication of sigmoid. Being analog makes the gates differentiable which is an advantage in back propagation.

RNNs have been able to make a significant improvement in contrast to other machine learning techniques due to their ability to learn and carry out complicated transformations of data its ability maintaining long term as well as short term dependencies. RNN contains interplay of Reasoning, Attention and Memory, commonly referred to as the ‘RAM model’ in Deep Neural Networks.

Humans are capable of understanding what is the information that they have to consider in order to answer correctly for a question asked by another person machines are lack of these capability. Injecting attention into a question and answering system will make the model answer in such a way that human does.

Overtime, different attention models have shown promising results as well. Researches have been done on how complex sequences with long-range structure can be generated with LSTM RNNs. [6]

In addition to that, Neural Machine Translation is another approach to machine translation, which attempts to build and train a single large neural network that reads and outputs a correct translation instead of having small sub-components that are tuned separately.[7] Most machine translators are encoder-decoder models where the encoder neural network reads and encodes a sentence to a fixed length vector and the decoder output the translation from the encoded vector. The major issue with this system is that the neural network should be able to compress all the important information in the sentence to a fixed-length vector. Thus, dealing with longer sentences becomes difficult. The distinguishing feature of this approach is that it does not encode the entire input sentence into a fixed length vector. Instead, it encodes the input into a sequence of vectors and choses a subset of these vectors adaptively while decoding.[7] This removes the potential challenge in dealing with long sentences and assists in retaining the necessary information.

A method has been proposed, utilizing a local attention-based model for Abstractive Sentence Summarization which up to date remains a challenge for Natural Language Processing. This model focuses on sentence-level summarization rather than using extracted potions of the sentences to prepare a condensed version. It uses a neural language model with an input encoder that learns a latent soft alignment over the input text to help inform the summary.

There are attention-based models introduced for Speech Recognition[8], Handwriting synthesis and Image Caption Generation as well.[9]

2.2. Identification and Significance of the Problem

Neural Networks are only capable of understanding numerically represented inputs. Word embedding becomes an essential part of any QA system because of this. As the context and the question are exclusively in a textual representation effective vector representation is co-related to the overall accuracy of the system. There are several approaches in word embedding that is used currently. Out of these the neural network approach stands to be the most promising. In this research I would be looking into the existing algorithms such as word2vec and glOVE and modify these concepts to create an algorithm based on the skipgram model.

The significance of the research is that we will be improving question preprocessing, Reasoning, Attention and Memory of our deep neural network which are vital and technically challenging tasks to carry out.

Though algorithms based on the word2vec model an improved version of the skipgram model is effective and used widely they have several fallbacks. These algorithms are highly reliant on their vocabulary so inability to handle unknown or out-of-vocabulary words is a significant issue. If the model has not encountered a word before it would be unable to interpret it and create a vector. Which then results in assigning a random vector that is far from ideal.

Another fallback of word2vec is that there are no shared representation at sub-word levels. As word2vec forms an individual vector for every word even though two words are morphologically similar the model would not be able to deduce the relationships between them.

Significant improvements in word embedding model would result in semantically rich vector representations of both context and query. This would increase the reasoning of the neural net as The details of the vector representation would be sufficient for the net to deduce and answer questions accurately.

2.3. Technical objectives

According to the explanations given in previous parts "NEURAL REASONING MODEL FOR QUESTION ANSWERING" project is based on dnn (Deep Neural Network) where the preprocessing, attention, reasoning and memory layers of dnn are enriched to produce better results than existing state-of-art techniques. This part of the report will demonstrate the Software and Hardware requirements of the research which will be used for performing technical tasks in order to improve the above mentioned for components.

2.3.1 Specific software requirements

The most renowned and accepted programming language for the machine learning problems is Python. Python is widely used across the world for solving machine learning problems for its readability and wide range of machine learning libraries. Our research project will be using Python as the main programming language.

TensorFlow is an open source artificial intelligence library, using data flow graphs to build models. It allows developers to create large-scale neural networks with many layers and TensorFlow allows distribution of computation across different computers, as well as multiple CPUs and GPUs within a single machine. TensorFlow provides a Python API so we will be using TensorFlow for our deep learning research.

Keras is a high-level neural network API, helping lead the way to the commoditization of deep learning and artificial intelligence. It runs on top of a number of lower-level libraries, used as backend, including TensorFlow and Theano. Keras code is portable, meaning that you can implement a neural network in Keras using Theano as a backend and then specify the backend to subsequently run on TensorFlow, and no further changes would be required to your code. Theano is a Python library for fast numerical computation that can be run on the CPU or GPU.It is a key foundational library for Deep Learning in Python that you can use directly to create Deep Learning models or wrapper libraries that greatly simplify the process. And this research will be using Keras and Theano as key software resources.

Overall Software Stack :

- Python
- Tensorflow
- Keras.
- Theano.

2.4. Detail Design

Preprocessing is the initial step in QA which enables question and context understood. The outputs of this components will be responsible for outputs of all other components. Machines do not understand natural language similar to humans. Thus, it is necessary to interpret the text inputted in an accurate machine-understandable form yet preserving the context. Vector Representations have been a popular [1][2].

The main objective of this component would be provide the neural network with the most effective vector representation which the rest of the components can utilize in order to provide the user with the most accurate answer. In order to do so, the Question type, the expected answer type and the question focus should be identified[3]. There multiple several techniques such as Word Embedding, Syntactic Analysis[4] and WH-question analysis to achieve this.[5]

In the proposed approach we will be focusing on Word Embedding. Word Embedding is used to map words or phrases to the corresponding vector. It allows words with similar meaning to have similar representation. It is done through algorithms such as Word2Vec, GloVe. The two learning models introduced to learn the word embedding in Word2Vec are Continuous Bag of Words model which predicts the current word based on its context and Continuous and Continuous Skip-Gram model which predicts the surrounding words given a current word.[7] But in this approach it is proposed to use GloVe algorithm which is extension to Word2Vec. It combines both the global statistics of matrix factorization techniques like Latent Semantic Analysis with the local context-based learning in Word2Vec.[7] The approach followed may change based on trial and error.

2.5. Sources for test data & analysis

The amount of data that we create is growing at a staggering pace and keeping track of it becomes more difficult. Accordingly, it is important that we use techniques to make the information that we need accessible. A question answer system is concerned with automatically answering questions posed by humans naturally, which would return the most specific or direct answer rather than returning a set of documents relating like in search engines. There are two broad types of question answering systems that could be built upon. One is Closed-domain question answering system, which basically deals with questions under a specific domain (bAbi – a closed domain dataset provided by Facebook). The other type is Open domain question answering system, which is concerned with question answering about almost any kind of question posed, therefore, it is required to use the information provided in the dataset as well as any additional available knowledge(WIKIQA - an open domain dataset). However, for the purpose of the research, we have settled our center of interest towards an open domain question answering system. This does not necessary mean that it will answer every open domain question. We would focus on answering general questions with an enhanced Deep Neural Network than state-of-the-art systems.

We plan to build a QA system using deep learning to help interpret a question posed and provide a suitable answer from a given context. We will facilitate the QA system to be built under any domain, complying with certain conditions to increase performance. The precision and the weight of reliability of the provided answers will majorly depend upon the accuracy of the context. However the goal in this research is to emphasize that by using deep learning techniques by combining Reasoning, Attention and Memory mechanisms, we are able to reduce some of the complexities and barriers that are present at the moment and provide an innovative product that utilizes the platform.

2.6. Anticipated Benefits

As we demonstrated above the latest improvements in deep learning methodologies made significant impact in natural language QAS for producing human like answers for a given question. This lead to creation of many QAS in open domain and closed domain with their significant changes and specialties. Through "NEURAL REASONING MODEL FOR QUESTION ANSWERING" research we tend to improve four major components of a dnn which are preprocessing, attention, reasoning and memory in order to produce a human like answers for a given question captivating state-of-art techniques. So that the user will be highly benefited with the answer than misleading with an incorrect answer. This will make user more enthusiastic to us our QAS system more and more.

By taking place of this research component of Corpus Preprocessing will add up as a contribution to the body of knowledge under deep neural networks. As the research carrying out currently on this component, have understood the major drawbacks of the existing corpus preprocessing techniques as explained in the Literature Review chapter. So this research is focused on overcoming such drawbacks and coming up with a good corpus preprocessing technique with the use of existing techniques and algorithms along with new enhancements.

By using attention mechanism on the facts, we refine the internal representation of facts based on the information (relevance) from the question. By aggregating relevant information into the final representation, we provide necessary information to the answer selection layer, to predict the answer to the question. This form of multi-hop "search" (on facts) allows the model to learn to perform some sophisticated reasoning for solving certain challenging tasks. This will ignore the irrelevant parts from the answer automatically so he answers will be more related to the question.

Humans are capable of reasoning about entities and relationships intuitively. However, machines possess this capability up to some extent only. Machines would have to utilize the previously acquired knowledge to do so. This would allow machines to reason about their entities and relations from unstructured data. Solving this would open infinite possibilities for Artificial

Intelligence. Through the reasoning the model will gain artificial intelligent which makes answers more human interactive.

Apart from technical benefits identifying an approach that deviates from current approaches and if the evaluation of the approach proves to have outperformed current state-of-art approaches this enables future work in the Intelligent Question and Answering domain to adapt and evolve the improvement of RAM model we have used.

3. PROJECT PLAN OR SCHEDULE

This figure below shows the project plan we follow as a team. By using this Gantt chart it will be easier to schedule tasks and workload within the team. This increases the efficiency of the team as well.

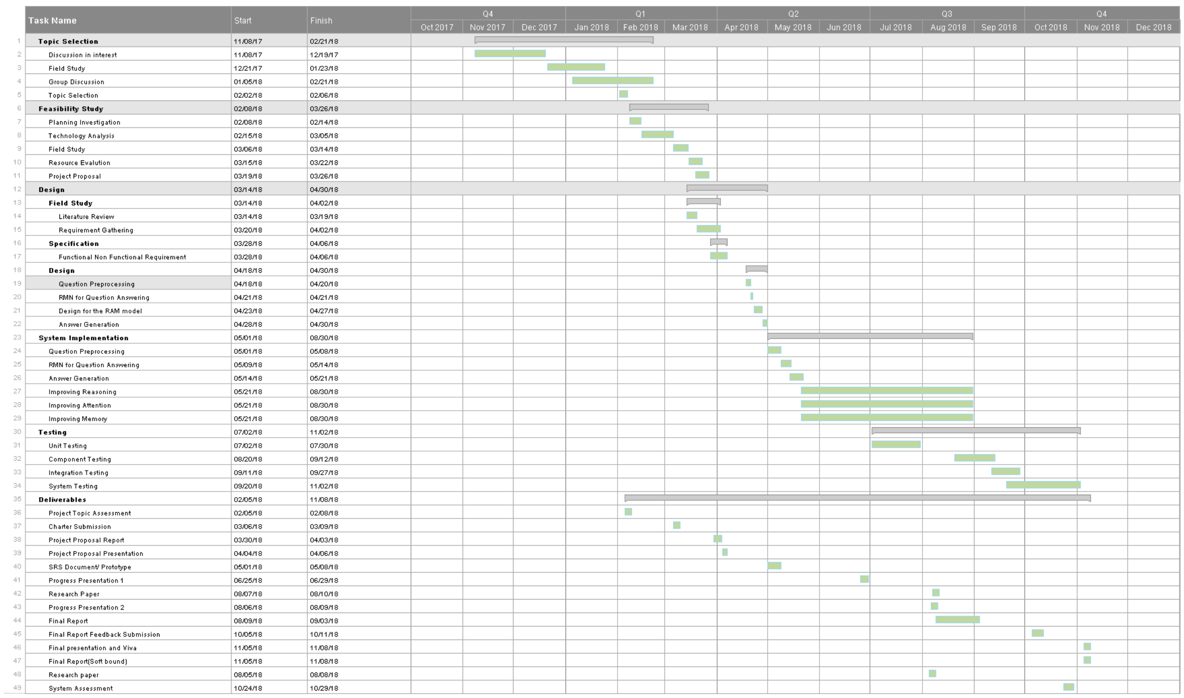

Figure 3.0 Gantt Chart of the Component

4. RESEARCH CONSTRAINTS

One of the main constraints that we have to face is the suitability of the dataset for our model. Although, we are planning to use the TriviaQA[29] dataset initially, we may have to change the dataset based on Trial and Error.

Another major constraint is that the training process is time consuming and in order to speed up this process it is required to use high end servers with GPU processing capabilities. Using these resources are costly and infeasible at the development stage. Therefore the development iterations might be time consuming.

Another challenge is that there must be continuous monitoring of the training process to avoid overfitting the model to the given dataset. The challenge is to balance overfitting and underfitting when training the model. This can be challenging sometimes with a neural net that is as complicated as RNN's.

The lack of expertise and online help in the context of Information Extraction using DNN is another constraint in this research component. Since we are focusing more on an implementation rather than a highly mathematical study, online help would be highly beneficial. Therefore this is another research constraint that we have to face.

It is clear that like every research there are several research constraints that should be tackled and dealt with appropriately in order to make this research a success.

5. SPECIFIED DELIVERABLES

As the main outcome of the project we aim to deliver a learning domain specific QA system with better performance through enhanced reasoning, memory and attention.

6. REFERENCES

[1] Aclweb.org, 2018. [Online]. Available: http://www.aclweb.org/anthology/P08-1028. [Accessed: 01- Apr- 2018].

[2] Homepages.inf.ed.ac.uk, 2018. [Online]. Available: http://homepages.inf.ed.ac.uk/s0453356/composition.pdf. [Accessed: 01- Apr- 2018].

[3] A. Ben Abacha and P. Zweigenbaum, "MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies", 2018. .

[4] Linguistics.ucla.edu, 2018. [Online]. Available: http://linguistics.ucla.edu/people/stabler/isat.pdf. [Accessed: 01- Apr- 2018].

[5] Aclweb.org, 2018. [Online]. Available: http://www.aclweb.org/anthology/D15-1237. [Accessed: 01- Apr- 2018].

[6] N. Gupta, "Artificial Neural Network", Network and Complex Systems, vol. 3, no. 1,pp. 24-28, 2013.

[7] J. Brownlee, "What Are Word Embeddings for Text? - Machine Learning Mastery", Machine Learning Mastery, 2018. [Online]. Available: https://machinelearningmastery.com/what-are-word-embeddings/. [Accessed: 01- Apr- 2018].

[8] Arxiv.org, 2018. [Online]. Available: https://arxiv.org/pdf/1703.04617.pdf. [Accessed: 01- Apr- 2018].

[9] 2018. [Online]. Available: https://www.researchgate.net/publication/320658588_Reinforced_Memory_Network_for_Question_Answering. [Accessed: 01- Apr- 2018].

[10] J. Hanlon and J. Hanlon, "How To Solve The Memory Challenges Of Deep Neural Networks - TOPBOTS", TOPBOTS, 2018. [Online]. Available: https://www.topbots.com/how-solve-memory-challenges-deep-learning-neural-networks-graphcore/. [Accessed: 01- Apr- 2018].

[11] "RAM Workshop", Thespermwhale.com, 2018. [Online]. Available: http://www.thespermwhale.com/jaseweston/ram/. [Accessed: 01- Apr- 2018].

[12] Memory Networks. Jason Weston, Sumit Chopra, Antoine Bordes. International Conference on Representation Learning, 2015

[13] Teaching Machines to Read and Comprehend. Karl Moritz Hermann et. al. arXiv Pre-Print, 2015.

[14] Large-scale Simple Question Answering with Memory Networks. Antoine Bordes, Nicolas Usunier, Sumit Chopra, Jason Weston. arXiv Pre-Print, 2015.

[15] L. Bottou, "From machine learning to machine reasoning", Machine Learning, vol. 94, no. 2, pp. 133-149, 2013.

[16] Arxiv.org, 2018. [Online]. Available: https://arxiv.org/pdf/1410.3916.pdf. [Accessed: 01- Apr- 2018].

[17] Proceedings.mlr.press, 2018. [Online]. Available: http://proceedings.mlr.press/v48/kumar16.pdf. [Accessed: 01- Apr- 2018].

[18] Arxiv.org, 2018. [Online]. Available: https://arxiv.org/pdf/1803.03067.pdf. [Accessed: 01- Apr- 2018].

[19] "How does an attention mechanism work in deep learning? - Quora", Quora.com, 2018. [Online]. Available: https://www.quora.com/How-does-an-attention-mechanism-work-in-deep-learning. [Accessed: 01- Apr- 2018].

[20] "How does attention model work using LSTM? - Quora", Quora.com, 2018. [Online]. Available: https://www.quora.com/How-does-attention-model-work-using-LSTM. [Accessed: 01- Apr- 2018].

[21] Aclweb.org, 2018. [Online]. Available: http://www.aclweb.org/anthology/C16-1290. [Accessed: 01- Apr- 2018].

[22] S. Dwivedi and V. Singh, "Research and Reviews in Question Answering System", 2018. .

[23] "Machine Learning for Question Answering | Machine Learning Group", Mlg.ulb.ac.be, 2018. [Online]. Available: http://mlg.ulb.ac.be/questionanswering. [Accessed: 01- Apr- 2018].

[24] "Intro to Deep Learning for Question Answering", Slideshare.net, 2018. [Online]. Available: https://www.slideshare.net/TraianRebedea/intro-to-deep-learning-for-auestion-answering. [Accessed: 01- Apr- 2018].

[25] "Optimizing Memory Consumption in Deep Learning — mxnet documentation", Mxnet.incubator.apache.org, 2018. [Online]. Available: https://mxnet.incubator.apache.org/architecture/note_memory.html. [Accessed: 01- Apr- 2018].

[26] Aclweb.org, 2018. [Online]. Available: http://www.aclweb.org/anthology/P15-2116. [Accessed: 01- Apr- 2018].

[27] Thespermwhale.com, 2018. [Online]. Available: http://www.thespermwhale.com/jaseweston/ram/papers/paper_18.pdf. [Accessed: 01- Apr- 2018].

[28] "A Brief Overview of Attention Mechanism – Synced – Medium", Medium, 2018. [Online]. Available: https://medium.com/@Synced/a-brief-overview-of-attention-mechanism-13c578ba9129. [Accessed: 01- Apr- 2018].

[29] "TriviaQA", Nlp.cs.washington.edu, 2018. [Online]. Available: http://nlp.cs.washington.edu/triviaqa/. [Accessed: 01- Apr- 2018].

[30] J. Bian, B. Gao and T. Liu, "Knowledge-Powered Deep Learning for Word Embedding", 2018. .

[31] S. Dwivedi and V. Singh, "Research and Reviews in Question Answering System", 2018. .

[32] "Understanding LSTM Networks -- colah's blog", Colah.github.io, 2018. [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/. [Accessed: 01- Apr- 2018].

[33] Cs224d.stanford.edu, 2018. [Online]. Available: https://cs224d.stanford.edu/reports/StrohMathur.pdf. [Accessed: 01- Apr- 2018].

[34] J. Hanlon and J. Hanlon, "How To Solve The Memory Challenges Of Deep Neural Networks - TOPBOTS", TOPBOTS, 2018. [Online]. Available: https://www.topbots.com/how-solve-memory-challenges-deep-learning-neural-networks-graphcore/. [Accessed: 01- Apr- 2018].

[35] Krizhevsky, A., Sutskever, I. & Hinton, “G. ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems”, 2012. [Accessed: 03- Apr- 2018]

[36] A. Eitel, J. T. Springenberg, L. Spinello, M. Riedmiller and W. Burgard, "Multimodal deep learning for robust RGB-D object recognition," *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, 2015, pp. 681-687. [Accessed: 03- Apr- 2018]

[37] WildML. (2018). *Attention and Memory in Deep Learning and NLP*. [online]Available at: http://www.wildml.com/2016/01/attention-and-memory-in-deep-learning-and-nlp/ [Accessed 4 Apr. 2018].

[38] Siegelmann, H. T. and Sontag, E. D. (1995). "On the computational power of neural nets". Journal of computer and system sciences.

[39] Alex Graves, Greg Wayne, Ivo Danihelka. (2014). 'Neural Turing Machines". [Online]. Available: https://arxiv.org/pdf/1410.5401v2.pdf. [Accessed 4 Apr. 2018].

[40] Armand Joulin Facebook, Tomas Mikolov (2015). "Inferring Algorithmic Patterns with Stack-Augmented Recurrent Nets". [Online]. Available: https://arxiv.org/pdf/1503.01007.pdf. [Accessed 4 Apr. 2018].

[41] Wojciech Zaremba, Ilya Sutskever (2016). "REINFORCEMENT LEARNING NEURAL TURING MACHINES". [Online]. Available: https://arxiv.org/pdf/1505.00521.pdf. [Accessed 4 Apr. 2018].

[42] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus (2015). "End-To-End Memory Networks". [Online]. Available: https://arxiv.org/pdf/1503.08895.pdf. [Accessed 4 Apr. 2018].

[43] Alex Graves (2014). "Generating Sequences With Recurrent Neural Networks". [Online]. Available: https://arxiv.org/pdf/1308.0850.pdf. [Accessed 4 Apr. 2018].

[44] Dzmitry Bahdanau, KyungHyun Cho, Yoshua Bengio. (2016) "NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE". [Online]. Available: https://arxiv.org/pdf/1308.0850.pdf. [Accessed 4 Apr. 2018].

[45] Jan Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Yoshua Bengio. (2015) "Attention-Based Models for Speech Recognition". [Online]. Available: https://arxiv.org/pdf/1308.0850.pdf. [Accessed 4 Apr. 2018].

[46] Kelvin Xu, Jimmy Lei Ba, Ryan Kiros, Kyunghyun Cho , Aaron Courville, Ruslan Salakhutdinov, Richard S. Zemel, Yoshua Bengio. (2016) “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention”. [Online]. Available: https://arxiv.org/pdf/1502.03044.pdf . [Accessed 4 Apr. 2018].

[47] Towards Data Science. (2018). *A Gentle Introduction To Neural Networks Series — Part 1*. [online] Available at: https://towardsdatascience.com/a-gentle-introduction-to-neural-networks-series-part-1-2b90b87795bc [Accessed 13 May 2018].

[48] freeCodeCamp. (2018). *Want to know how Deep Learning works? Here’s a quick guide for everyone.*. [online] Available at: https://medium.freecodecamp.org/want-to-know-how-deep-learning-works-heres-a-quick-guide-for-everyone-1aedeca88076 [Accessed 13 May 2018].

[49] Chris V. Nicholson, S. (2018). *A Beginner's Guide to Recurrent Networks and LSTMs - Deeplearning4j: Open-source, Distributed Deep Learning for the JVM*. [online] Deeplearning4j.org. Available at: https://deeplearning4j.org/lstm.html#recurrent [Accessed 13 May 2018].