International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Analysis on the Language Independent and Dependent Aspects of Deep Learning based Question Answering Systems

R. Poonguzhali, K. Lakshmi

Abstract: Natural languages are ambiguous and computers are not capable of understanding the natural languages in the way people really understand them. Natural Language Processing (NLP) is concerned with the development of computational models based on the aspects of human language processing. Question Answering (QA) system is a field of Natural Language Processing that provides precise answer for the user question which is given in natural language. In this work, a MemN2N model based question answering system is implemented and its performance is evaluated with a complex question answering tasks using bAbI dataset of three different language text corpuses. The scope of this work is to understand the language independent and dependant aspects of a deep learning network. For this, we will study the performance of the deep learning network by training and testing it with different kinds of question answering tasks with different languages and also try to understand the difference in performance with respect to the languages

Keywords : NLP, QA, Deep learning, MemNN, Memory Networks, MemN2N, End to End Memory Networks, RNN, LSTM, GRU, bAbI Tasks.

I. INTRODUCTION

Natural Language has very rich forms and structure which is very ambiguous. There are five main categories into which language ambiguities fall: syntactic, lexical, semantic, referential and pragmatic. Understanding of natural languages would be much more difficult which requires understanding both the word and the context to give the specific message. The ambiguity and vague characteristics of languages make NLP, difficult for the machines to understand [22]. The earlier NLP techniques based on statistical and machine learning techniques could not be able to handle these categories of language ambiguities and also fail to achieve good results in a typical NLP task.

But, in recent days, the growth of deep learning networks and related technologies promising human level accuracy and efficiency in most of the NLP related tasks.

Revised Manuscript Received on April 30, 2020.

* Correspondence Author

R. Poonguzhali*, Computer Science and Engineering, Periyar Maniammai Institute of Science and Technology (Deemed to be University), Thanjavur, India. Email: poonguzhali079@gmail.com

K. Lakshmi, Computer Science and Engineering, Periyar Maniammai Institute of Science and Technology (Deemed to be University), Thanjavur, India. Email: lakshmi@pmu.edu

For example, most of the solutions and deep learning models for question answering systems are giving more than human level of accuracy. Deep learning networks are capable of providing a solution for a NLP task without the real understanding on the grammar and other aspects of the language. They simply learn anything from given examples.

Deep Learning network can process multiple languages at once and can learn language independent model from the given text corpus.

Memory networks are a new class model developed to solve the long term dependencies problem. In this work, we will try to study the language independent and dependant aspects of a deep learning network using the state of the art End to End memory network (MemN2N) model and evaluate its performance with different language text corpus.

The first QA systems were BASE BALL (1961) and LUNAR (1972). LUNAR and BASEBALL were good at answering the corresponding domains. Question Answering era started in 1999. Text Retrieval conference (TREC) made the challenge on Open domain question answering. Even though Current QA systems deal with simple factual questions, more system needed to answer complex questions. One such question is temporal question [7].

QA system can automatically and accurately answer the questions posed in natural language. The end user of this system is interested to receive direct answer to an information need. The sources of information for QA system are documents, audio, text, files or databases. The QA system may be broadly classified as Restricted Domain Question Answering (RDQA) and Open Domain Question Answering (ODQA). RDQA systems answer questions posed by users in a specific domain which rely on manually constructed data or some knowledge sources. ODQA focuses on answering questions apart from the subject domain which extracts from a large corpus of textual documents which may be semi-structured or unstructured [3].

NLP [20] is an Artificial Intelligence branch that makes the machines to read, understands and generates meaning from the human languages. This field of study mainly focuses on the interaction between computers and human languages. The process of understanding and manipulating language is extremely complex. This human-computer interaction enables real-world applications like automatic text summarization, machine translation, sentiment analysis, topic extraction, named entity recognition, parts-of-speech tagging, relationship extraction, stemming, and more. NLP is widely known for machine translation and automated question answering. NLP problems are named as an AI-complete problem.

It is very difficult to make computers to think and solve problem as human does. Strong-AI system performs activities more than human [1].

Deep Learning [21] is an artificial intelligence technology that replicates the structure and functions of the human brain in computing data and finding patterns during decision making. Deep learning can be considered as a branch of machine learning in Artificial Intelligence (AI) that has networks which has the potential to learn from unsupervised data which is also known as deep neural learning or deep neural network. Deep Neural Networks (DNNs) are networks where each layer can do complex operations such as representation and abstraction that make sense of images, sound, and text.

It is a field which is based on learning and improving on its own. So far, neural networks were restricted by computing power and thus were limited in complexity. Nevertheless recent developments in technologies have evolved larger and convenient neural networks make computers to learn and react to complex situations faster than humans. Recently, Deep learning models attain good accuracy. Deep learning has given enormously promising results for various tasks in natural language understanding, especially on the subject of topic classification, sentiment analysis, and question answering and language translation. It can be used to solve any pattern recognition problem without human intervention.

This work intends to evaluate the performance of the state of the art End to End memory network (MemN2N) model with different QA tasks and with different language/text corpuses.

In this work, a MemN2N model based question answering system is implemented and their performance is evaluated with a complex question answering tasks using bAbI dataset. We will study the performance of training and testing with three different text corpuses with suitable metrics and try to find the difference in performance with respect to the languages in which the task is presented to the network.

The paper is organized as follows. Section 2 explains state-of-the-art Memory Network models. Section 3 explains about the Modeling of general QA system. In section 4, the results are discussed based on the three different language/text Corpus From bAbI Dataset.

II. THE MEMORY NETWORKS AND END TO END MEMORY NETWORKS

A. The Memory Networks (MemNN)

In [6] the authors present a new learning model called memory networks with inference components and long-term memory component. This model used both the components effectively.

There are four components I, G, O and R and a memory mi. The work flow of this model is described as follows:

- The input sentence is converted into an internal feature representation I (x).
- The Generalization component adds new elements to the memory mi with the new input: m_i = G (m_i, I (x), m), ∀i.
- Using the new input and the memory, compute the output features o: o = O (I (x), m).
- Lastly, get the final response: r = R(o).

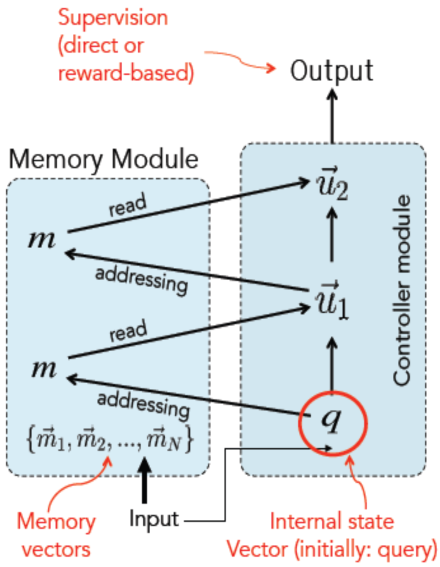

Figure 1: MemNN Model

Implemention of MemNN model

I (input): The input is converted into vectors using bag-of-word-embeddings x.

G (generalization): This component stores x in next available slot m_N.

O (output): The output feature component loops over the memories k=1 or 2 times:
1st loop max: finds best match mi with x.
2nd loop max: finds best match m_J with (x, m_i).
The output O is represented with (x, m_i, m_J).

R (response): The reponse component o ranks all words in the dictionary for the best match and returns best single word.

Training :

Training is performed with a margin ranking loss and stochastic gradient descent (SGD). We minimize over model parameters U_O and U_R:

Minimize:

$$\sum_{\bar{f} \neq m_{O_1}} \max(0, \gamma - s_O(x, m_{O_1}) + s_O(x, \bar{f})) +$$
$$\sum_{\bar{f}' \neq m_{O_2}} \max(0, \gamma - s_O([x, m_{O_1}], m_{O_2}]) + s_O([x, m_{O_1}], \bar{f}'])) +$$
$$\sum_{\bar{r} \neq r} \max(0, \gamma - s_R([x, m_{O_1}, m_{O_2}], r) + s_R([x, m_{O_1}, m_{O_2}], \bar{r}])) + \qquad (1)$$

Where,

S_O is the matching function for the Output component.
S_R is the matching function for the Response component.
x is the input question.
m_{O1} is the first true supporting fact.
m_{O2} is the first second supporting fact.
r is the response

True facts and responses should have higher scores than all other facts and responses by a given margin.

$\bar{\mathbf{f}}$, $\bar{\mathbf{f}}'$ and $\bar{\mathrm{r}}$ are the predicted choices, and γ is the margin. The sample is generated $\bar{\mathbf{f}}$, $\bar{\mathbf{f}}'$, $\bar{\mathrm{r}}$ at every level of SGD.

If an RNN is used instead of R component of MemNN, the last term was replaced with the standard log likelihood used in a language modeling task.

B. The End to End Memory Network (MemN2N)

Single layer:

The input context $\{x_1.....x_n\}$ are converted into two embedding matrices A and C with dimension d and vocabulary size V (d × V). Those matrices are represented using two memory cells input *mi* and output c_i. The question q is also embedded with u to determine the correlation between both which is computed by taking the inner product followed by a softmax function:

$$p_i = Soft\max(u^T m_i) \tag{2}$$

where $$Soft\max(z_i) = e^{z_i} / \sum_j e^{z_j} \tag{3}$$

and *p* is a probability vector of the inputs.

Hence, the response o from the output memory cell is computed by a weighted sum of the transformed inputs c_i, and the probability vector:

$$0 = \sum_i p_i c_i \tag{4}$$

Finally the answer is predicted by

$$\hat{a} = Soft\max(W(o+u)) \tag{5}$$

Multiple Layers:

The model can be extended to add the memory layers to handle several K hop operations described below:

The memory layer k+1 th hop takes the input and the output from layer k as follows:

$$u^{k+1} = u^k + o^k \tag{6}$$

Finally, the answer prediction for the question q combines top memory layer input and output:

$$\hat{a} = Soft\max(Wu^{k+1}) = Soft\max(w(o^k + u^k)) \tag{7}$$

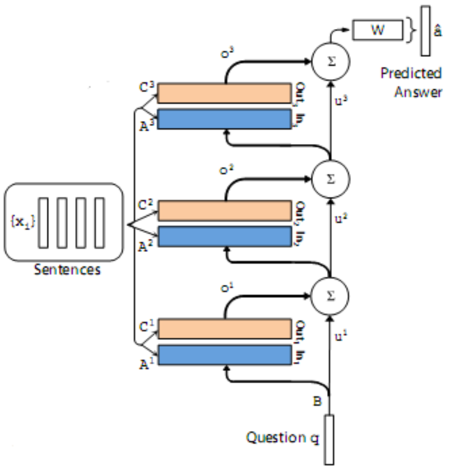

Figure 2: A 3 Layer MemN2N Model.

A three-layer version of memory model is shown in the above figure. The soft max operations are performed in each layer in MemN2N Model.

C. The QA bAbI tasks for NLP Research

In [4], the authors proposed a set of tasks to test the reasoning skills. It is a set of 20 synthetic tasks. Each task contains training data and test data. In this dataset, a given QA task includes set of statements, question and set of facts. The answer is given during the training period and predicted at the testing time. The answer may be a single word or set of words based on the tasks.

The tasks in bAbI dataset are given below:

ID	Task	Class name
1	Basic factoid QA with single supporting fact	WhereIsActor
2	Factoid QA with two supporting facts	WhereIsObject
3	Factoid QA with three supporting facts	WhereWasObject
4	Two argument relations: subject vs. object	IsDir
5	Three argument relations	WhoWhatGave
6	Yes/No questions	IsActorThere
7	Counting	Counting
8	Lists/Sets	Listing
9	Simple Negation	Negation
10	Indefinite Knowledge	Indefinite
11	Basic coreference	BasicCoreference
12	Conjunction	Conjunction
13	Compound coreference	CompoundCoreference
14	Time manipulation	Time
15	Basic deduction	Deduction
16	Basic induction	Induction
17	Positional reasoning	PositionalReasoning
18	Reasoning about size	Size
19	Path finding	PathFinding
20	Reasoning about agent's motivation	Motivations

III. MODELLING A QA SYSTEM

A. Training a Typical QA System

The following diagram shows the typical process of training a questing answering system. Using a word-vector dictionary, the training story texts, question texts and their corresponding answers texts will be vectorized and a deep learning network will be trained with the vectorized training data.

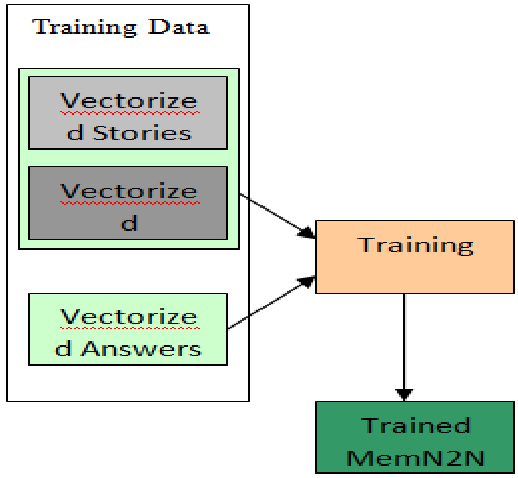

Figure 3: Training MemN2N

B. Testing a Typical System

The following diagram shows the typical process of testing or validating a questing answering system

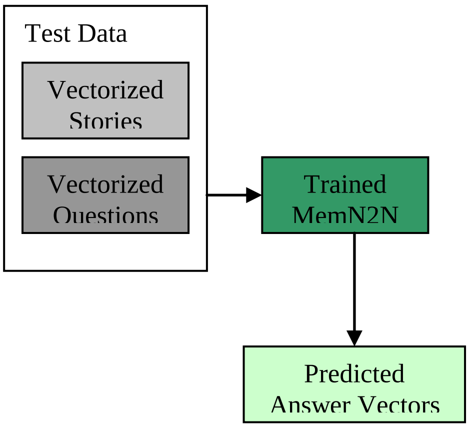

Figure 4: Testing MemN2N

The test story texts, and question texts will be vectorized and fed in to the trained network and the network will predict the possible answer vectors. The actual answer test from the vectorized answers will be created using reverse lookup in the word-vector dictionary.

C. About Keras

Keras is a open source neural network API which is capable of running on top of Tensor Flow, CNTK, or Theano. It is written in python. Keras can *be used if deep leaning library is needed. It create fast prototype model with the advantage of user-friendliness, modularity, and extensibility. It also supports both CNN and RNN. It executed on CPU and GPU. It is designed to make fast implementation with deep neural networks.*

Keras executes the model in the way we created with defined layers or using multiple input-output models. Keras also run our model with loss and optimizer functions and training process with fit function. Keras doesn't handle low level computations which are handled by backend engine. This engine performs all the computations at back end level such as tensor products using the libraries like Tensor Flow or Theno**.** Tensorflow is the "backend engine" that is default but it can be changed in the configuration.

D. Tensor Flow

Tensor Flow is an open source software library that operates at large scale and in heterogeneous environments. Its flexible architecture maps the nodes of a dataflow graph across many machines in a cluster, and within a machine across a variety of platforms (CPUs, GPUs, TPUs). It is implemented by the Google Brain team researchers and engineers within Google's AI organization. It speed up research in machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains. Tensor Flow develops several applications, with a focus on training and inference on deep neural networks.
Tensor-Flow is designed to run faster on GPUs than CPUs using cuda library built on NVidia card.

IV. RESULTS AND DISCUSSION

About the three different language/text Corpus Used From bAbI Dataset*:*

The bAbI dataset contains the following directories of dataset like English, Hindi, and Shuffled letters. The samples are given in both 1000 and 10,000.In each directory, we can find 20 QA tasks in a specific language.
The scope of this evaluation is to study the performance of the deep learning network with same kind of QA tasks that were represented in three different languages English, Hindi, and a Shuffled version of English which will not be human readable form.

The following three boxes show the same kind of QA task (Task-ID 19 of bAbI) and its prediction result (predicted answer).

Results of Task-19 in English:

the garden is west of the hallway
the kitchen is west of the garden
the garden is north of the bathroom
the bedroom is east of the bathroom
the hallway is west of the office

Question: how do you go from the bathroom to the hallway

Original Answer: n,e
Prediction Answer: n,e

Results of Task-19 in Hindi:

galiyara sayanakaksh ki uttar disha mein hai
galiyara bagichey ki poorav disha mein hai
rasoi ghar bagichey ki paschim disha mein hai
gusalkhana bagichey ki uttar disha mein hai
daftar bagichey ki dakshin disha mein hai
Question: gusalkhana sey galiyara tak jaaney ka kya rasta hai
Original Answer: dakshin,poorav
Prediction Answer: dakshin,poorav

Results of Task-19 in Shuffled Letter Representation:

rzh rtxahm el jhlq bs qzh ztuujti
rzh keqozhm el jhlq bs qzh rtxahm
rzh rtxahm el mbxqz bs qzh ptqzxbby
rzh phaxbby el htlq bs qzh ptqzxbby
rzh ztuujti el jhlq bs qzh bsseoh
Question: dbj ab ibd rb sxby qzh ptqzxbby qb qzh ztuujti
Original Answer: m,h
Prediction Answer: j,j

Overall Results

Training Parameters:

Total Training Samples: 1000
Total Testing/Validation Samples: 1000
No Epochs: 100
Training Batch Size: 32

The following table shows the overall results of Task-19 from three different language/text corpuses

Retrieval Number: F4816049620/2020©BEIESP
DOI: 10.35940/ijitee.F4816.049620
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 1: The Results with The bAbI tasks ID: 19

Language	No Trainable Parameters	Time Taken for Training (s)	Accuracy	Loss
English	14800	211	0.388	1763.17
Hindi	16800	209	0.435	1762.74
Shuffled	15200	203	0.378	1807.68

The following table shows the overall results of Task-16 from three different language/text corpuses.

Table 2: The Results with The bAbI tasks ID: 16

Language	No Trainable Parameters	Time Taken for Training (s)	Accuracy	Loss
English	10800	182	0.715	627.12
Hindi	14000	236	0.688	732.00
Shuffled	10800	175	0.751	616.38

The following table shows the overall results of Task-17 from three different language/text corpuses.

Table 3: The Results with The bAbI tasks ID: 17

Language	No Trainable Parameters	Time Taken for Training (s)	Accuracy	Loss
English	8400	114	0.808	463.04
Hindi	8400	101	0.725	583.14
Shuffled	9200	114	0.771	524.96

The following bar chart shows the number of trainable parameters used for the MemN2N network during solving three different QA tasks in three different languages/texts. As shown in this chart, the number of trainable parameters depends on the language in which the task is represented. In Task19 and Task16, the Hindi version of the same task required more parameters than the English version. In Task17, the parameters are almost equal in all language representations.

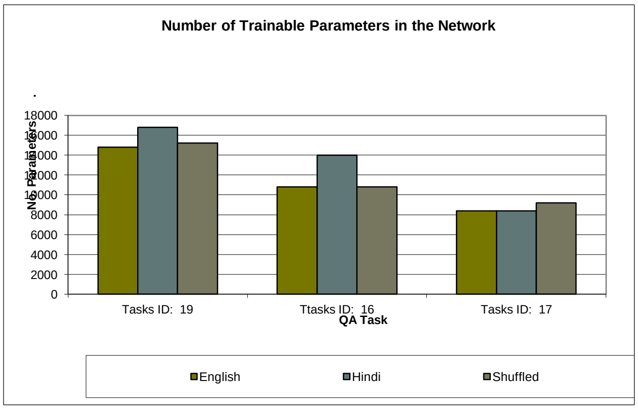

Figure 5: The Comparison of No. of Training Parameters

The following bar chart shows the training performance (in terms of training time) of MemN2N network for solving three different QA tasks in three different languages/texts.

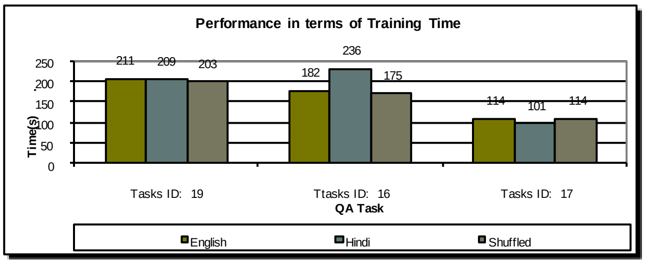

Figure 6: The Comparison of Training Time

The following bar chart shows the performance of MemN2N network in terms of accuracy in solving three different QA tasks in three different languages/texts. As shown in this chart, the language in which the task is represented has some positive and negative impact on performance. Hindi representation of the same task has some significant impact on the performance. If we carefully notice the performance of English and Shuffled version of the tasks, we can say that the accuracy in these two cases is almost equal. This may be because, the Shuffled version is nothing but character shuffled version of English text.

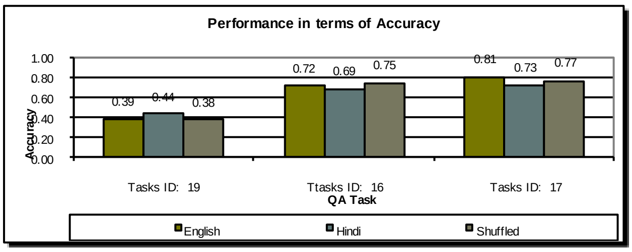

Figure 7: The Comparison of Accuracy of Different Languages on Different Tasks

The following bar chart shows the performance of MemN2N network in terms of MSE/loss/score for solving three different QA tasks in three different languages/texts. As shown in this chart, here also, the language in which the task is represented has some positive and negative impact on performance. Hindi representation of the same task has some impact on the performance in the case of task16 and 17.

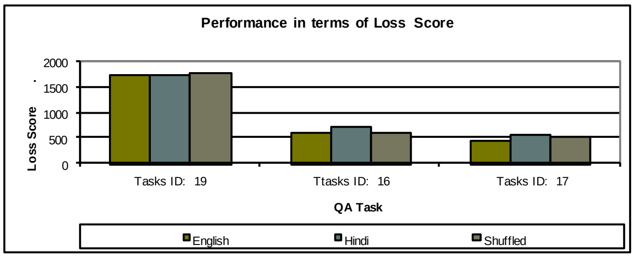

Figure 8: The Comparison of Loss in Different Models

V. CONCLUSION

We have tested the performance of MemN2N based questing answering system with three different language/text corpuses. As we expected, the language in which the task is represented has some positive and negative impact on performance of the deep learning network as well as the size and complexity of the network itself. We realized that, the Hindi representation of the same task has significant impact on the performance in terms of different metrics. If we carefully notice the performance of English and Shuffled version of the tasks, we can say that the accuracy in these two cases is almost equal in most of the cases. This may be because, the Shuffled version is nothing but character shuffled version of English text and almost equal in other aspects.

As we know that one deep learning network designed for a QA task is capable of handling more than one language text without any modification in its design. But, the results of this work prove that the performance of the system will depend on the language in which the task is represented. This may be because of the inherent differences in languages and their grammar. As we know, without any knowledge on such grammar or syntax of a language, deep learning network is capable of learning it, but certainly the performance will depend on grammar and syntax related aspects of the language.

There are possibilities of designing a deep learning network based QA system which will give improved performance irrespective of the nature of the language in which the problem is represented. Further, there will be several ways to improve the MemN2N network to achieve improved performance in complex QA tasks. Our future works will address these issues.

ACKNOWLEDGEMENT

The authors thank the Management of Periyar Maniammai Institute of Science & Technology (Deemed to be University) for providing needed resources and moral support to complete the work successfully.

REFERENCES

1. Deeksha Dwivedi, Mahendra Singh Sagar, "Overview of Natural Language Processing", 4th International Conference on System Modeling & Advancement in Research Trends (SMART), 2015.
2. Tom Young, Devamanyu Hazarika, Soujanya Poria, Erik Cambria,"Recent Trends in Deep Learning Based Natural Language Processing", arXiv:1708.02709v8 [cs.CL], 25 Nov 2018.
3. Kolomiyets Oleksander and Marie-Francine Moens, A survey on question answering technology from an information retrieval perspective, in: Journal of Information Sciences volume 181, 2011.
4. Jason Weston, Antoine Bordes, Sumit Chopra, Tomas Mikolov, Alexander M. Rush, "Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks", arXiv:1502.05698v10 [cs.AI] 31 Dec 2015, (conference paper at ICLR 2016).
5. Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus, "End-To-End Memory Networks", arXiv:1503.08895v5 [cs.NE] 24 Nov 2015.
6. Jason Weston, Sumit Chopra & Antoine Bordes, "Memory Networks", arXiv:1410.3916v11 [cs.AI], 29 Nov 2015, (conference paper at ICLR 2015).
7. Poonguzhali, K Lakshmi, "Temporal Question Answering System: A Survey", JETIR June 2019, Volume 6, Issue 6, ISSN-2349-5162.
8. Yashvardhan Sharmaa, Sahil Guptaa, "Deep Learning Approaches for Question Answering System", International Conference on Computational Intelligence and Data Science (ICCIDS 2018).
9. K.S.D. Ishwari, A.K.R.R.Aneeze, S.Sudheesan, H.J.D.A. Karunaratne, A. Nugaliyadde, Y.Mallawarrachchi, "Advances in Natural Language Question Answering: A Review", arXiv:1904.05276 [cs.CL], Apr 2019.
10. Emily M. Bender, "Linguistically Na¨ive != Language Independent: Why NLP Needs Linguistic Typology", Proceedings of the EACL 2009 Workshop on the Interaction between Linguistics and Computational Linguistics, pages 26-32, Athens, Greece, 30 March, 2009.
11. Kodra and E. Kajo, "Question Answering Systems: A Review on Present Developments, Challenges and Trends", International Journal of Advanced Computer Science and Applications, vol. 8, no. 9, 2017
12. E. Brill, J. Lin, M. Banko, S. Dumais and A. Ng, "Data-Intensive Question Answering", Trec.nist.gov, 2018.
13. H. Madabushi and M. Lee, "High Accuracy Rule-based Question Classification using Question Syntax and Semantics", Aclweb.org, 2018.
14. E. Riloff and M. Thelen, "A Rule-based Question Answering System for Reading Comprehension Tests", 2018.
15. S. Humphrey and A. Brownea, "Comparing a Rule Based vs. Statistical System for Automatic Categorization of MEDLINE Documents According to Biomedical Specialty", 2018.
16. S. K. Dwivedia and V. Singh, "Research and reviews in question answering system," in Proceedings of International Conference on Computational Intelligence: Modeling Techniques and Applications, 2013, pp. 417 - 424 .
17. D. Cohn, Z. Ghahramani and M. Jordan, "Active Learning with Statistical Models", Journal of Artificial Intelligence Research, vol.4, 1996.
18. X. Li and D. Roth, "Learning Question Classifiers", Dl.acm.org, 2018.
19. Raghuvanshi, A., & Chase, P., "Dynamic Memory Networks for Question Answering".
20. Ramesh Sharda et.el, "Business Intelligence and Analytics: Systems for Decision Support", Copyright© 2015, 2011, 2007 by Pearson Education, Inc., ISBN 10: 0-13-305090-4, ISBN 13: 978-0-13-305090-5
21. Elvis, "Deep Learning for NLP: An Overview of Recent Trends", https://medium.com/dair-ai
22. Finlay, Janet, and Alan Dix. An introduction to Artificial Intelligence. London: UCL Press, 1996.
23. R Poonguzhali, Dr K Lakshmi, " Evaluating the Performance of Recurrent Neural Network based Question Answering System with Easy and Complex bAbI QA Tasks", International Journal of Advanced Science and Technology, Vol. 29, No. 5s, ISSN: 2005 – 4238 pp. 1389-1402, (2020).
24. R Poonguzhali, K Lakshmi, "Analysis on the Performance of Some Standard Deep Learning Network Models for Question Answering Task", Manuscript accepted for publication.
25. R Poonguzhali, K Lakshmi, "Evaluating the Performance of Keras Implementation of MemNN Model for Simple and Complex Question Answering Tasks", In: Test Engineering and Management, Volume 82, ISSN: 0193 - 4120 pp. 9620-9629, (2020).
26. R Poonguzhali, K Lakshmi, "Evaluating the Performance of MemN2N Model for a Complex Question Answering Task", Manuscript submitted for publication.

AUTHORS PROFILE

R.Poonguzhali has completed B.E (CSE) in Government College of Engineering (GCE), Tirunelveli and obtained her M.E in Software Engineering from Periyar Maniammai College of Technology for Women - PMCTW (Affiliated to Anna University), Tamil Nadu. Currently she is pursuing Ph.D in the Department of Computer Science and Engineering from Periyar Maniammai Institute of Science and Technology (Deemed to be University), Thanjavur, Tamil Nadu. She has 20 years of experience in teaching field. Her research involves Natural Language Processing, Machine Learning, and Deep Learning. She has membership in Professional bodies like ISTE and CSI.

Dr.K.Lakshmi has received her Ph.D. degree from Anna University, College of Engineering - Guindy. She has been doing research in the area of Machine Learning, Text Mining, Data Mining, Computational Linguistics and Image Processing. She has published many research papers in the mentioned areas. She has a working experience of more than 20 years. She had taken up many administrative positions like Headship of Department, Deanship of Faculty etc.